





# THEORETICAL PHYSICS

by

**GEORG JOOS**

*Professor of Experimental Physics at the Technische  
Hochschule, Munich*

With the Collaboration of

**IRA M. FREEMAN**

*Professor of Physics, Rutgers University*

**THIRD EDITION**

London • **BLACKIE & SON LIMITED** • Glasgow

198  
—  
C.

*[Faint, illegible stamp]*

*First published 1934*  
*Reprinted 1940, 1941, 1942, 1943, 1944 (twice),*  
*1946, 1947 (twice)*  
*Second edition 1951*  
*Reprinted 1951, 1953, 1954, 1956, 1957*  
*Third edition 1958*  
*Reprinted 1958, 1959, 1960, 1961*



ALLAMA IQBAL LIBRARY



45816

BT

SA 2

30.1  
7417



## PREFACE TO THE SECOND EDITION

FROM its inception, the aim of this book has been to bring the reader to an intermediate level of attainment in the main branches of theoretical physics from which he may be able to proceed, with the help of special literature, to the field of research. Experience shows that it is often the first contact with a complex problem that presents the major difficulty, and it is here that help is most urgently needed. Nevertheless, it is necessary to make a certain selection from the many topics that seem to offer promise of development. For example, there is the question of whether an extensive treatment of the Hamilton-Jacobi mechanics, which at the time of the first edition (1932) formed the basis of the old atom theory, is still advisable. The decision to retain it is founded on the conviction that only in this way is the compelling and logical evolution of atomic physics in the form of wave mechanics clearly seen.

Ensuing revisions of the book have made possible the recasting of the presentation of many topics. The subject of nuclear physics required, of course, the most extensive changes. In this connexion it is felt that a concise survey of the main lines of cosmic ray research is now in order, since knowledge in this field has progressed far beyond the stage of mere speculation. In this brief presentation much of the great mass of experimental material has been omitted, and it is hoped that specialists in this field will not be too critical of what is apparently a somewhat oversimplified picture. Geometric optics has been given more space and the theory of the top has been modernized on the basis of a treatment suggested by Professor Bauersfeld.

The range of subject matter has been extended by the inclusion of selected topics in what might be called "applied theoretical physics". To have included these items in the respective chapters dealing with these subjects would have interrupted the continuity of development, and so they have been grouped in a separate Part of the book. In response to numerous suggestions, a Mathematical Addendum on the properties of Bessel Functions and Spherical Harmonics has been

prepared. This material has been placed intentionally at the end of the work rather than in the Mathematical Introduction, in as much as it demands a somewhat higher order of computational skill on the part of the reader.

Other changes include: (a) introduction of the M.K.S. system, chiefly in the formulas of macroscopic electromagnetism (there is no reason to change the familiar numerical relations in atomic physics); (b) addition of a considerable number of new exercises; (c) revision of the numerical values of physical constants; and (d) extension of the list of references for further study. It is hoped that these alterations will increase the usefulness of the work in its dual role of text and reference book.

G. J.  
I. M. F.

BOSTON, MASS., AND  
NEW BRUNSWICK, N.J.,  
U.S.A.  
*April, 1950.*

## PREFACE TO THE THIRD EDITION

In this revision there are two additional chapters: (1) Phenomenological Theory of Superconductivity, (2) Theory of Elastomers; and the chapter on Nuclear Physics is considerably modified. A section on Fundamentals of the Matrix Calculus has been added to the first chapter and one on The Role of Lattice Defects in Dielectric Crystals to Chapter XLI. The Table of Physical Constants has also been revised.

# CONTENTS

*Sections marked with an asterisk are somewhat more difficult  
and may be omitted in the first perusal of the work*

## PART I

### MATHEMATICAL INTRODUCTION

#### CHAPTER I

#### Vector Analysis

	Page
1. The Concept of a Vector . . . . .	7
2. Addition and Subtraction of Vectors; Multiplication of a Vector by a Scalar . . . . .	8
3. The Scalar Product of two Vectors . . . . .	10
4. The Vector Product of two Vectors. The directed Plane Area as a Vector . . . . .	12
5. Multiple Products . . . . .	15
6. Differentiation of a Vector with respect to a Scalar; Application to the Theory of Space Curves . . . . .	17
7. Space Derivatives of a Scalar Quantity . . . . .	20
8. The Concept of Divergence and Gauss's Theorem . . . . .	23
9. The Curl of a Vector, and Stokes's Theorem . . . . .	25
10. The Operator $\nabla$ . . . . .	30
11. Calculation of the Gradient in a Vector Field; Fundamental Principles of Tensor Analysis . . . . .	32
12. Calculation of more complicated Vector Differential Expressions with the help of the Nabla Operator . . . . .	37
13. Differential Vector Operations in Curvilinear Orthogonal Co-ordinates .	40



	Page
14. Degeneration of the Vector Differential Operations at Surfaces of Discontinuity in the Field - - - - -	43
15. Fundamentals of the Matrix Calculus - - - - -	46

## CHAPTER II

## Mathematical Representation of Periodic Phenomena; Theory of Vibrations and Waves

1. Simple Harmonic Vibrations - - - - -	50
2. Representation of more complicated Periodic Phenomena by Series of Harmonic Terms. Fourier Series. The Fourier Integral - -	54
3. Modulated Vibrations and Beats - - - - -	57
4. Combination of Vibrations along different Axes. Lissajous' Figures -	58
5. The Propagation of Periodic Disturbances in the form of Waves - -	61
6. Combination of several Waves having the same Direction of Propagation; Linearly and Elliptically Polarized Waves; Group Velocity - - - - -	64
7. Combination of Waves having the same Frequency but different Directions of Propagation. Standing Waves - - - - -	67

## CHAPTER III

## Selected Topics in the Theory of Functions of a Complex Variable

1. Conformal Mapping of one Plane on another - - - - -	69
2. The Cauchy-Riemann Conditions and the Differential Equation of Laplace - - - - -	71
3. Line Integrals in the Gauss Plane; the Cauchy Integral Theorem -	72

## CHAPTER IV

## The Fundamental Problem of the Calculus of Variations and its Solution

1. Statement of the Problem of the Calculus of Variations - - -	75
2. Derivation of the Euler-Lagrange Differential Equation - - -	75

## PART II

## MECHANICS

### CHAPTER V

### The Mechanics of a Single Particle

	Page
1. The Fundamental Concepts of Kinematics - - - - -	81
2. Newton's Second Law of Motion - - - - -	83
3. Time Integral and Path Integral of the Force. Work and Energy -	85
4. Conservative Forces; Potential - - - - -	87
5. Central Forces; the Law of Areas - - - - -	88
6. Gravitational Forces; Planetary Motion - - - - -	90
7. Quasi-elastic Forces and Harmonic Vibrations - - - - -	93
8. Harmonic Vibrations with Frictional Resistance - - - - -	94
9. Forced Vibrations; Resonance - - - - -	96
10. Non-harmonic Vibrations; Sudden Changes of Amplitude - - -	100
11. Mechanics of a Constrained Particle. The Simple Pendulum - -	103

### CHAPTER VI

### General Theorems on the Mechanics of Systems of Particles

1. Theorem concerning the Motion of the Centre of Mass - - - -	108
2. Angular Momentum of a System of Particles - - - - -	110
3. Total Energy of a System of Particles - - - - -	111
4. The Principle of Virtual Displacements, D'Alembert's Principle and the Lagrangian Equations of the First Kind - - - - -	113
5. The Lagrangian Equations of the Second Kind for Arbitrary Co- ordinates (Generalized Co-ordinates) - - - - -	118
6. Generalized Momentum Co-ordinates. Hamilton's Equations - -	121
7. Hamilton's Principle - - - - -	123
8. Canonical Transformations - - - - -	124
9. Cyclic Variables. The Hamilton-Jacobi Differential Equation - -	126
10. Periodic and Multiply Periodic Systems. Angle Variables; the Angle Variables of the Keplerian Motion - - - - -	128

## CHAPTER VII

**The Mechanics of Rigid Bodies**

	Page
1. Selected Topics in the Kinematics of Rigid Bodies . . . . .	138
2. General Statics and Dynamics of Rigid Bodies. Equivalence of Systems of Forces acting upon Rigid Bodies . . . . .	143
3. Rotation of a Rigid Body about a Fixed Axis. Moment of Inertia and its Calculation . . . . .	146
4. Motion of a Rigid Body about a Fixed Point. Elements of the Theory of the Top . . . . .	151

## CHAPTER VIII

**Elasticity: The Mechanics of Deformable Solids**

1. The Geometry of Small Displacements . . . . .	161
2. State of Stress of a Body under Strain . . . . .	167
3. The Conditions of Equilibrium of an Elastic Body . . . . .	169
4. Relations between the Strain Tensor and the Stress Tensor . . . . .	171
5. Energy of Elastically Deformed Bodies; Elastic Potential . . . . .	175
6. Elementary Treatment of the Bending of a Cantilever Beam . . . . .	176
7. Waves in Unbounded Elastic Media (Seismic Waves.) Longitudinal Waves in Bars . . . . .	180
8. Transverse Vibration of Stretched Strings and Membranes . . . . .	183

## CHAPTER IX

**The Mechanics of Liquids and Gases  
(Hydro- and Aero-Mechanics)**

1. Equilibrium of Fluid Bodies (Hydrostatics) . . . . .	191
2. The Fundamental Hydrodynamical Equations . . . . .	194
3. Irrotational Flow . . . . .	197
4. General Theorems concerning Vortex and Circulation . . . . .	203
5. Plane Circulatory Motion . . . . .	207



# CONTENTS

	xi
	Page
6. Undulatory Propagation of Disturbances in Fluids (Sound Waves) . . . . .	212
7. Hydrodynamics of Viscous Fluids . . . . .	214
8. Surface Tension of Liquids . . . . .	222

## CHAPTER X

### Relativistic Mechanics

1. Space and Time in Newtonian Mechanics . . . . .	228
2. Inertial Frames. The Galilean Transformation . . . . .	229
3. Accelerated Frames of Reference. Free Fall on the Rotating Earth . . . . .	231
4. Moving Frames of Reference in Acoustics. The Doppler Effect . . . . .	235
5. Moving Frames of Reference in Optics. The Michelson-Morley Experiment . . . . .	237
6. The Relativistic Conception of Space and Time. The Lorentz Transformation . . . . .	240
7. Immediate Consequences of the Lorentz Transformation . . . . .	242
8. Geometric Representation of the Lorentz Transformation. The Four-dimensional World. Calculation with World-vectors . . . . .	247
9. Newton's Second Law in the Theory of Relativity. The Variability of Mass and the Inertia of Energy . . . . .	252
10. Fundamental Concepts of the Generalized Theory of Relativity . . . . .	258

## PART III

### FIELD THEORY OF ELECTROMAGNETIC AND OPTICAL PHENOMENA

## CHAPTER XI

### The Electrostatic Field in a Vacuum (or in Air)

1. Definitions . . . . .	263
2. Electrical Charge (Quantity of Electricity) as the Source of Flux . . . . .	264
3. The Electrostatic Potential . . . . .	267
4. Simple Examples of the Electrostatic Field in a Vacuum (or in Air) . . . . .	269

## CHAPTER XII

**The Electrostatic Field in Dielectric Media**

	Page
1. Formal Introduction of the Concepts " Dielectric Displacement " and " Free Charge ". Boundary Conditions at the Surface of Separation of Two Dielectrics - - - - -	277
2. Polarization of Dielectrics - - - - -	279
3. Simple Examples of the Electrostatic Field in Dielectrics - - -	282

## CHAPTER XIII

**Energy and Ponderomotive Forces in the Electrostatic Field**

1. Potential Energy of Systems of Charges in a given Field - - -	285
2. Total Energy of the Electrostatic Field - - - - -	286
3. Forces and Equilibrium in the Electrostatic Field. Theory of the Manometer Method for the Electrical Susceptibility of a Liquid - - -	290

## CHAPTER XIV

**Stationary Electric Fields and Steady Currents**

1. Ohm's Law - - - - -	293
2. Generation of Heat in a steady Electric Field - - - - -	295

## CHAPTER XV

**The Magnetostatic Field**

1. Comparison of Electrostatic and Magnetostatic Fields - - -	297
2. Calculation of the Magnetostatic Field accompanying a given Distribution of Electric Currents in a Vacuum - - - - -	300
3. Calculation of the Magnetic Field accompanying Electric Currents when Ferromagnetic Materials are present - - - - -	305
4. Ponderomotive Forces on Conductors in a Magnetic Field - - -	308

CHAPTER XVI

**Slowly Varying (Quasi-stationary) Fields**

	Page
1. The Law of Induction. Maxwell's Equations . . . . .	311
2. Self and Mutual Induction . . . . .	313
3. Steady Alternating-current Circuits . . . . .	316
4. Non-stationary States (Transient Phenomena) in Alternating-current Circuits . . . . .	319
5. Resistance and Inductance of Wires for Alternating Currents. Skin Effect . . . . .	324

CHAPTER XVII

**Rapidly Alternating Electromagnetic Fields:  
I—Propagation in Homogeneous Isotropic Media**

1. The Electrical Analogue of the Law of Induction . . . . .	327
2. The Wave Equation for the Propagation of Fields in Dielectrics . . . . .	328
3. The Poynting Vector of Energy Flow . . . . .	331
4. Propagation of Electromagnetic Waves in Conduction Media . . . . .	333
5. Hertz's Solution of the Field Equations. The Hertzian Oscillator . . . . .	335

CHAPTER XVIII

**Electromagnetic Waves: II—Phenomena in  
Two Adjoining Media**

1. Unified Rigorous Derivation of the Field Equations and of the Boundary Conditions . . . . .	343
2. Waves in the Boundary Layer. The Ground Wave . . . . .	345
3. Consequence of the Boundary Conditions for Insulating Media. The Optical Laws of Reflection and Refraction . . . . .	349
4. Polarization and Intensity Relationships for Reflection and Refraction. Fresnel's Formulæ . . . . .	350
5. Total Reflection . . . . .	354
6. Absorbing Media. The Optics of Metals . . . . .	357

## CHAPTER XIX

## Electromagnetic Waves: III—Propagation in Anisotropic Media. The Optics of Crystals

	Page
1. The Field Equations for Anisotropic Bodies . . . . .	360
2. Plane Electromagnetic Waves in Anisotropic Media . . . . .	363
3. Normal Surface and Wave Surface. The Optical Axes . . . . .	368
4. Refraction of Plane Waves at the Surface of an Anisotropic Medium . . . . .	371

## CHAPTER XX

## Electromagnetic Waves: IV—The Theory of Diffraction

1. The General Diffraction Problem and Attempts to solve it. Kirchoff's Formula . . . . .	376
2. Reciprocal Theorems of the Theory of Diffraction. Classification of Diffraction Phenomena . . . . .	379
3. Fraunhofer Diffraction by a Slit and by One-, Two- and Three-dimensional Gratings . . . . .	382
4. Fresnel Diffraction Phenomena at a Slit and at a Circular Aperture. Zone Plates . . . . .	390

## CHAPTER XXI

## The Elements of Geometrical Optics and of Interference Optics

1. The Fundamentals of Geometrical Optics. Laws of Fermat and of Malus . . . . .	395
2. The Properties of Collinear Projection . . . . .	397
3. The Practical Problem of Image Formation. Abbe's Sine Law. General Path of an Elementary Pencil from a Point Source . . . . .	403
4. The Resolving Power of Optical Systems . . . . .	409
5. The Fundamentals of Interference Optics. Interference Fringes . . . . .	410

## PART IV

THE THEORY OF ELECTRICITY. II. THE ATOMISTIC  
NATURE OF ELECTRICAL PHENOMENA

## CHAPTER XXII

**Electrolytic Conduction**

	Page
1. The Fundamental Phenomena of Electrolytic Conduction and their Interpretation . . . . .	417
2. Dependence of Electrolytic Conductivity upon Concentration. The Theory of Debye-Hückel and Onsager . . . . .	421

## CHAPTER XXIII

**The Conduction of Electricity in Gases**

1. Direct Determination of the Elementary Electrical Charge by the Millikan Oil Drop Method . . . . .	425
2. The Nature of the Cathode Ray Particles. The Electron . . . . .	426
3. Survey of the Possible Methods of Generating Carriers of Electricity in Gases . . . . .	431
4. The Separately-sustained Electrical Discharge. Spark Discharge . . . . .	433
5. Self-maintaining Discharge; Glow and Arc Discharge . . . . .	437
6. The Origin of Cathode, Canal, and Positive Rays. The Mass Spectrograph . . . . .	440

## CHAPTER XXIV

**The Fundamental Ideas of the Theory of  
Metallic Conduction**

1. Electrons as Carriers of Current in Metals . . . . .	444
2. Derivation of Ohm's Law for Metals . . . . .	446
3. Conduction of Heat in Metals; the Law of Wiedemann and Franz . . . . .	447
4. Objections to the Theory developed above. The Electron Theory of Pauli and Sommerfeld . . . . .	449



## CHAPTER XXV

## Electron Theory of the Dielectric Constant, Index of Refraction and Magnetic Permeability

	Page
1. The Origin of Electrical and Magnetic Polarization . . . . .	450
2. Theory of Dielectric Polarization, Optical Index of Refraction and Dispersion . . . . .	451
3. Pielectric Susceptibility . . . . .	455
4. Paramagnetic, Ferromagnetic and Antiferromagnetic Susceptibility .	458
5. Magnetism Induced by Revolving Electrons. The Magnetomechanical Parallelism. Theory of Diamagnetic Susceptibility . . . . .	463

## CHAPTER XXVI

## Phenomenological Theory of Superconductivity

1. The Fundamental Equations . . . . .	466
2. Steady Fields . . . . .	468
3. Optical Behaviour of Superconductors . . . . .	469

## CHAPTER XXVII

## The Electrodynamics of Moving Bodies

1. Electromagnetic Induction in Moving Bodies from the Standpoint of the Electron Theory . . . . .	471
2. Magnetic Effects of Moving Charges . . . . .	474
3. Propagation of Electromagnetic Waves in Moving Media . . . . .	477
4. Relativistic-invariant Form of the Electromagnetic Equations . . . . .	480

## PART V

## THE THEORY OF HEAT. PHENOMENOLOGICAL PART

## CHAPTER XXVIII

## Theory of the Conduction of Heat

	Page
1. Definition of Temperature, Quantity of Heat, Thermal Capacity and Specific Heat - - - - -	487
2. The Differential Equation of Heat Conduction; Initial and Boundary Conditions - - - - -	489
3. A Simple Example of the Integration of the Equation of Heat Conduction: Penetration of the Daily and Yearly Temperature Variations into the Interior of the Earth - - - - -	490

## CHAPTER XXIX

## The Equation of State of Thermodynamic Systems

1. Definition of the Thermodynamic Variables and the Relationships between them - - - - -	492
2. The Equation of State of an Ideal Gas - - - - -	494
3. The Equation of State of a Real Gas - - - - -	497

## CHAPTER XXX

## The First Law of Thermodynamics: The Conservation of Energy

1. Formulation of the First Law - - - - -	501
2. Specific Heat at Constant Volume and at Constant Pressure. The Energy Function of a Gas - - - - -	503
3. Adiabatic Change - - - - -	506
4. Application of the First Law to Thermochemistry - - - - -	510

## CHAPTER XXXI

## The Second Law of Thermodynamics: The Law of Entropy

	Page
1. The Carnot Cycle and the Ideal Heat Engine - - - - -	514
2. Formulation and Interpretation of the Second Law - - - - -	518
3. Conditions of Equilibrium for Systems under Various Conditions. Thermodynamic Potentials - - - - -	524
4. Connexion between the Internal Energy and the Equation of State -	527
5. Electrocaloric and Magnetocaloric Phenomena - - - - -	528

## CHAPTER XXXII

## Application of the Second Law to the Calculation of the Equilibrium of Thermodynamic Systems

1. Gibbs' Phase Rule: a General Theorem on the Maximum Number of Possible Phases - - - - -	531
2. The Vapour Pressure Curve and the Melting-point Curve - - - - -	534
3. Chemical Equilibrium in a Mixture of Ideal Gases. The Law of Mass Action - - - - -	539
4. Chemical Equilibrium in a System consisting of Dilute Solutions and Ideal Gases - - - - -	543
5. Thermodynamic Equilibrium of Dilute Solutions of Strong Electrolytes	547

## CHAPTER XXXIII

## The Nernst Heat Theorem

1. Free Energy as a Measure of Chemical Affinity; Determination of this Quantity for the Galvanic Cell - - - - -	553
2. Formulation of the Nernst Heat Theorem - - - - -	555
3. Consequences for the Specific Heats and Temperature Coefficients. The Chemical Constant of a Gas - - - - -	557
4. Unattainability of the Absolute Zero - - - - -	559



## PART VI

## THE THEORY OF HEAT. STATISTICAL PART

## CHAPTER XXXIV

**The Elementary Kinetic Theory of Matter**

	Page
1. Bernoulli's Formula; Boyle's Law - - - - -	563
2. Number of Collisions and Mean Free Path for Real Gases - - -	565
3. Viscosity and Heat Conduction in Gases. Determination of Avogadro's Number and of the Size of the Molecules - - - - -	568
4. Derivation of some Properties of Crystals on the Molecular Theory -	573

## CHAPTER XXXV

**The Classical Statistics of Boltzmann**

1. Entropy and Probability - - - - -	579
2. Calculation of the most Probable Distribution of Density in an Ideal Gas	580
3. Representation of the Distribution of Position and Velocity by means of Phase Space. Liouville's Theorem - - - - -	584
4. The Maxwell-Boltzmann Energy Distribution - - - - -	586
5. Applications of the Maxwell-Boltzmann Energy Distribution - -	589
6. The Law of Equipartition and its Application to the Specific Heats -	595
7. Fluctuations. Transition from Microscopic to Macroscopic Motion -	598

## CHAPTER XXXVI

**The Classical Quantum Statistics**

1. The Quantum Condition for the Subdivision of the Phase Space of an Oscillator - - - - -	603
2. Temperature Variation of the Vibrational Contribution to the Specific Heats - - - - -	605
3. The Debye Theory of the Specific Heats of Solids - - - - -	608

## CHAPTER XXXVII

**The Theory of Thermal Radiation**

	Page
1. Analogy between a Gas and Radiation filling an Enclosure. Radiation Pressure - . . . . .	613
2. The Connexion between Energy Density and Surface Brightness - .	616
3. The Connexion between Emission and Absorption in Thermodynamic Equilibrium. Kirchhoff's Law - . . . . .	619
4. Planck's Law of Radiation - . . . . .	621

## CHAPTER XXXVIII

**The Bose-Einstein and the Fermi-Dirac Statistics**

1. Preliminary Remarks concerning the Duality of Waves and Corpuscles -	625
2. The Bose-Einstein Statistics - . . . . .	626
3. The Fermi-Dirac Statistics and its Application to Electrical Conductivity - . . . . .	631
4. Entropy Constant and Chemical Constant of an Ideal Monatomic Gas -	636

## PART VII

**THE STRUCTURE OF ATOMS AND MOLECULES  
AND THE THEORY OF SPECTRA**

## CHAPTER XXXIX

**The Mechanics of Simple Atomic Models**

1. Investigation of the Structure of the Atom; the Scattering of $\alpha$ -Particles - . . . . .	641
2. Conclusions from the Experiments on the Scattering of $\alpha$ -Particles -	645
3. The Bohr Model of the Hydrogen Atom - . . . . .	647
4. Consideration of the Motion of the Nucleus. The Spectrum of Hydrogen and the Spectrum of Ionized Helium - . . . . .	651

# CONTENTS

xxi

	Page
5. X-ray Spectra. Moseley's Law . . . . .	653
6. The Correspondence Principle . . . . .	655
7. The Atomic Spectra of the Alkalis, the Alkaline Earths and Similar Systems . . . . .	657
8. Perturbation of Electron Orbits by External Forces. Stark Effect and Zeeman Effect . . . . .	662
9. Difficulties arising in the Atomic Theory of Magnetism. Explanation in Terms of the Spinning Electron . . . . .	667
10. The Theory of Multiplets and of their Zeeman Effects. Quantum Theory of Paramagnetic Susceptibility . . . . .	669
11. The Structure of the Periodic System of the Elements. Pauli's Principle . . . . .	674
12. Elements of the Theory of Band Spectra . . . . .	680

## CHAPTER XL

### Critical Atomic Theory: The New Quantum Mechanics

1. The Wave and Particle Aspects of Light. Compton Effect and Raman Effect . . . . .	687
2. The Inexactness of Atomic Observations . . . . .	690
3. Matter Waves . . . . .	693
4. The Proper Values of the Wave Equation . . . . .	697
5. The Hydrogen Atom . . . . .	698
6. Rotation Spectra of Diatomic Molecules . . . . .	702
7. The Physical Meaning of the $\Psi$ Function. Intensity Relations for Spectral Lines . . . . .	703

## CHAPTER XLI

### Atomic Problems First Solved by the Wave Mechanics

1. The Theory of Perturbations in Wave Mechanics . . . . .	707
2. The Chemical Bond. Formation of the Hydrogen Molecule . . . . .	712
3. The Theory of Dispersion and of the Raman Effect in Wave Mechanics . . . . .	715
4. The Surmounting of Potential Barriers in the Wave Mechanics . . . . .	721
5. Energy Bands of Electrons in Metals . . . . .	725
6. The Role of Lattice Defects in Dielectric Crystals . . . . .	735

## CHAPTER XLII

**Nuclear Physics**

	Page
1. General Remarks on the Physics of the Atomic Nucleus . . . . .	739
2. The Hyperfine Structure of Spectral Lines as a Connecting Link between the Physics of the Outer Shells and that of the Nucleus . . . . .	743
3. Radioactivity . . . . .	745
4. Artificial Transformation and Excitation of the Nucleus . . . . .	747
5. The Neutron . . . . .	752
6. The Positron . . . . .	754
7. Artificial Radioactivity . . . . .	756
8. The Role of Protons and Neutrons in Nuclear Structure . . . . .	757
9. The Liquid Drop Model of the Nucleus . . . . .	763
10. The Fission of Uranium . . . . .	767
11. Liquid Drop Treated as a Fermi Gas . . . . .	769
12. Independent Particle and Shell Models of the Nucleus; Nuclear Moments . . . . .	772
13. Beta Transformations and the Neutrino Hypothesis . . . . .	777
14. Mesons . . . . .	778
15. Brief Survey of our Knowledge concerning Cosmic Rays . . . . .	780

## PART VIII

SELECTED TOPICS FROM SEVERAL FIELDS OF  
TECHNICAL IMPORTANCE

## CHAPTER XLIII

**Further Matters Connected with the  
Geometric Optics of Light and of Electrons**

1. Condition for Absence of Distortion and its Relation to the Sine Condition . . . . .	785
2. Focal Length of an Electron Lens . . . . .	787

CHAPTER XLIV

**Piezoelectricity and its Applications**

	Page
1. The Phenomenon of Piezoelectricity . . . . .	792
2. Application of Piezoelectricity to the Stabilization of Oscillating Circuits . . . . .	794

CHAPTER XLV

**Space-charge Effects in Gaseous Discharges**

1. Characteristic Curve of a Thermionic Tube . . . . .	799
2. Oscillations of a Plasma . . . . .	801

CHAPTER XLVI

**Theory of Elastomers**

1. Behaviour of High-polymer Chain Molecules . . . . .	803
2. A Model of Muscular Action . . . . .	805

MATHEMATICAL ADDENDUM

**Some Functions occurring in connection with the Wave Equation**

1. Generalization of $x!$ for any value of $x$ . . . . .	807
2. Bessel Functions of the First Kind . . . . .	810
3. Bessel Functions of the Second and Third Kinds. Asymptotic Values of Bessel Functions . . . . .	816
4. Spherical Harmonics . . . . .	819

SOLUTIONS OF THE EXERCISES . . . . .	825
TABLES . . . . .	865
REFERENCES FOR FURTHER STUDY . . . . .	868
INDEX . . . . .	873





# THEORETICAL PHYSICS

---

## INTRODUCTION

### THE PURPOSE AND METHODS OF THEORETICAL PHYSICS

Purely experimental physics confines its activities to the exhibition of natural phenomena and the careful description of what occurs. One might, for example, observe the luminous effects produced by the passage of an electric current through a rarefied gas, record what is observed, and stop at that point; but to-day a procedure of this kind is justly regarded as being unsatisfactory. As soon as we inquire into the *reasons* for the phenomena, we enter the domain of theory, which, on the basis of hypotheses admitting of more or less direct test, connects the observed phenomena and traces them back to single "pure" phenomena, thus bringing about a logical arrangement of an enormous amount of observational material.

In the given example of electrical discharge in gases, an "explanation"—i.e. a reference back to simple elements—requires theoretical concepts concerning the construction of the atom from electric charges and the mechanism of the production of charge-bearing particles. We thus find that the explanation of the phenomenon is far from simple, and that many effects operate simultaneously. Consequently, from the theoretical standpoint, we cannot look upon the phenomenon as "simple", however easy it is to produce experimentally. Theory alone can decide whether the experimental conditions are pure, i.e. simple from a theoretical point of view. In the present example, an experiment in which electrons of measurable uniform velocity collide with gas atoms may be regarded as simple, in spite of the considerable experimental difficulties, since on the basis of such experiments one can attempt to give an account of the more complex phenomena of the glow discharge.

While it is true that theory often sets difficult, if not impossible tasks for experiment, it does, on the other hand, often lighten the work of the experimenter by disclosing cogent relationships which make possible the indirect determination of inaccessible quantities, and thus render difficult measurements unnecessary. If it is required,

for example, to measure the reflecting power of a certain metal, the direct determination would require troublesome and withal not very accurate photometric measurements. However, if we are familiar with the theory of metallic reflection (p. 346), we will measure, instead, two angles—the principal angle of incidence and the principal azimuth—from which the value of the reflecting power is completely determined.

Whence does theory obtain its hypotheses concerning the connexions between individual physical quantities and events? In the last analysis, only from experiment. It is the art of the theorist to deduce from the data at hand, connexions of the most far-reaching significance, and to draw therefrom conclusions which suggest new experiments. There does not exist a theory which draws only upon itself, and which has no contact with experimental results. The number of possibilities is too large for even the greatest genius to be able to construct, by unaided intuition, a world-picture which agrees with experience. Because of their neglect of experiment, the scientists (natural philosophers) of ancient times exhausted themselves in speculations which only proved of lasting value where they were connected with observation, as in Astronomy and Mechanics. A train of thought of the nature of the well-known Hegelian dialogue \* seems quite absurd to us to-day, despite the fact that when contradictions between experiment and the consequences of a *well-founded* theory arise, we look for a mistake in the experimental procedure as often as we suspect an error in the calculations. In such a case, however, the theory has been set up to begin with so as to agree with observational facts. But while it is true that the first task of theory is to disclose relationships, an equally important part of its work is to formulate these relationships mathematically. Mathematics is the outstanding tool of the theorist. Its use represents a rationalization of thought, in that the process of deriving important conclusions from the initial hypotheses runs in the trustworthy channels of rules of calculation learned once for all. At the same time, one must not lose sight of the meaning of the computations. The beginner, especially, is inclined to ignore the physical meaning of his work, as contrasted with the purely formal calculations. The position occupied by mathematics in the field of theoretical physics implies that it is not the problem of the theoretical physicist to devise mathematical proofs. He must rely upon the soundness of the tool delivered to him by the mathematician. Even in a case where, occasionally, he must construct his own tool, he need not trouble about mathematical existence theorems, provided the result is sufficiently evident on physical grounds. Indeed, the strict requirements of pure mathematics often contradict the physical facts. For example, if density is defined as the *limiting* value of the ratio of mass to volume

\* "There can be only seven planets." "But this is contrary to fact." "So much the worse for the facts."



as smaller and smaller volumes are taken, the existence of atoms leads to meaningless density fluctuations, the values depending upon whether the volume element encloses an atomic nucleus or not. The "small quantities" prescribed by strict mathematics are indispensable in physics. Although we reckon with them as if they were "infinitely small" quantities, physical differentials have a not *too* small value. For such reasons, the ability to use the calculus as a tool is far more important to a physicist than a knowledge of strict concepts such as the limit mentioned above.

Theoretical and experimental physics cannot be separated, and a flawless, logically ordered picture of nature is obtained only by the application of both methods jointly. A rationally applied theory may have direct and immediate practical significance. If it is of importance to obtain the highest possible efficiency in a large-scale technical process, there is only one way to proceed: we must ascertain the most favourable conditions by exact calculations based on the theory; intuition is often deceptive. One need only recall, for example, the original pointed form of the airship, which seems to the great majority of people, if they are unfamiliar with hydrodynamics, to be the most effective form for cutting through the air.

Thus an exact knowledge of natural laws is sooner or later transmuted into economic advantages, which, as a sort of "premium on intelligence" \*, are gained by any society which organizes its industrial processes on rational principles.

\* The expression was coined by Schottky in the same connexion.



## PART I

### MATHEMATICAL INTRODUCTION

In this textbook the fundamental principles of the differential and integral calculus are assumed to be known. Any mathematics required beyond this point is collected in four preliminary chapters, in order to avoid later interruption of the physical train of thought by auxiliary mathematical discussions. A branch of mathematics of especial value to theoretical physics is *Vector Analysis*, which has been developed largely by physicists. Its concepts have immediate perceptual meaning, so that many physical laws show their full significance only when stated in the language of vectors. It is therefore quite a mistaken view to regard vector analysis merely as a computational shorthand.



# CHAPTER I

## VECTOR ANALYSIS

### 1. The Concept of a Vector.

Besides such magnitudes as temperature, mass, and so on, which are characterized by the assignment of a single number, and which are called *scalars*, there also occur in physics quantities which are not completely defined by the specification of a *single* number. The most important group of these is characterized by the fact that, besides the magnitude, the direction of the quantity must also be given. If, for example, we prescribe that an object, located by means of its centre of gravity  $P$ , be given a displacement

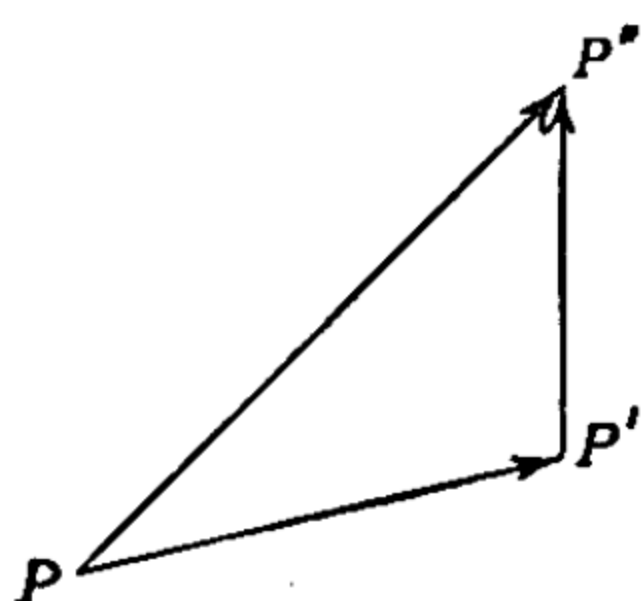


Fig. 1

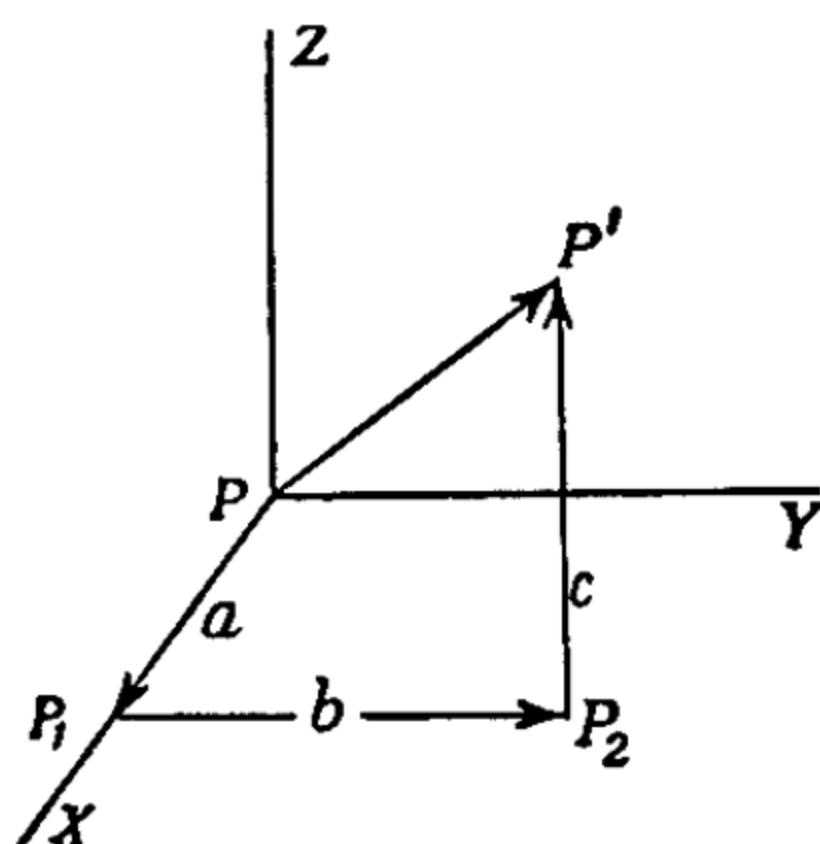


Fig. 2

of 2 cm., then the point  $P'$  occupied by the object after displacement may be anywhere on the surface of a sphere of radius 2 cm. with centre at  $P$ . The new position of the centre of gravity will not be determined uniquely unless we specify also the *direction* of the displacement in some manner. Two displacements of a point may be combined into a single equivalent displacement by a simple method (fig. 1). Instead of moving the point  $P$  to  $P'$  in a certain direction, and then displacing it in a new direction to  $P''$ , it might have been moved directly along the third side of the triangle  $PP'P''$  to  $P''$ . Since this process may be repeated, we can use it to characterize a displacement in still another way. We place  $P$  at the origin of a rectangular co-ordinate system, and move the point first along the  $x$ -axis to  $P_1$ , a distance  $a$ , then parallel to the  $y$ -axis a distance  $b$ , to  $P_2$ , and finally from here parallel to the  $z$ -axis to  $P'$ , a distance  $c$  (fig. 2).



Instead of this, we might have moved  $P$  at once in the direction  $PP'$  a distance

$$d = \sqrt{a^2 + b^2 + c^2}.$$

We see, then, that the displacement of  $P$  is uniquely determined by specifying the three numbers  $a$ ,  $b$  and  $c$ , which are called *components* of the displacement.

A large number of physical quantities follow the same law of combination as the displacement of a point, and may, therefore, also be defined by three numbers. Such quantities are called *vectors*. We may represent a vector diagrammatically by means of a directed line segment (arrow), whose length, on any convenient scale, is numerically equal to the magnitude of the physical quantity, and whose direction is that of the physical quantity. This geometric representation gives vector quantities an important advantage over other non-scalar quantities arising in physics, e.g. tensors, for which no such simple model exists.

Two vectors are said to be equal if they are identical with respect to both direction and magnitude. Since a pure displacement, i.e. a displacement without rotation, does not alter the direction, two vectors may still be equal, even when they lie in different (but parallel) lines. Physical quantities which are altered by pure displacement are therefore not immediately representable by vectors. To this class belongs, for example, the force acting on a rigid body capable of rotating about an axis: parallel displacement alters the lever arm of the force, and thus changes its effect. It is to be noted, further, that not every physical quantity which can be represented uniquely by a directed line segment may be treated as a vector; it is also necessary to investigate whether this quantity follows the *Law of Composition* obeyed by the displacement of a point. For instance, the rotation of a rigid body about an axis may be represented by an arrow whose direction is that of the axis of rotation, and whose length in centimetres is numerically equal to the rotation in degrees. Moreover, the direction of the arrow may be specified so that the rotation is clockwise when sighting in this direction. But we do *not* obtain a third rotation equivalent to the joint effect of two rotations by combining, as above, their representative arrows; the vectorial composition may be shown to be valid only for infinitesimal rotations.

In this book, vectors will be designated by heavy letters (bold-face or Clarendon type).

## 2. Addition and Subtraction of Vectors; Multiplication of a Vector by a Scalar.

The composition of two displacements of a point, described in § 1 (p. 7), has, as will be shown at once, all the characteristics of a

*summation.* In general, we define, therefore, as the sum of two vectors **A** and **B**, a vector **C** obtained by laying off the vector **B** from the terminal point of **A** and then drawing a vector from the initial point of **A** to the end point of **B** (fig. 3). This combination is written as a vector equation

$$\mathbf{C} = \mathbf{A} + \mathbf{B}. \quad . . . . . (1)$$

A vector equation is equivalent to *three* scalar equations, since, as was shown in § 1 (p. 8), a vector is determined by specifying its three components, and two vectors are equal only if the three components are respectively equal. { For if any vector is constructed as the combination of three vectors respectively equal to its components, two equal vectors can be obtained only if their corresponding components are identical.

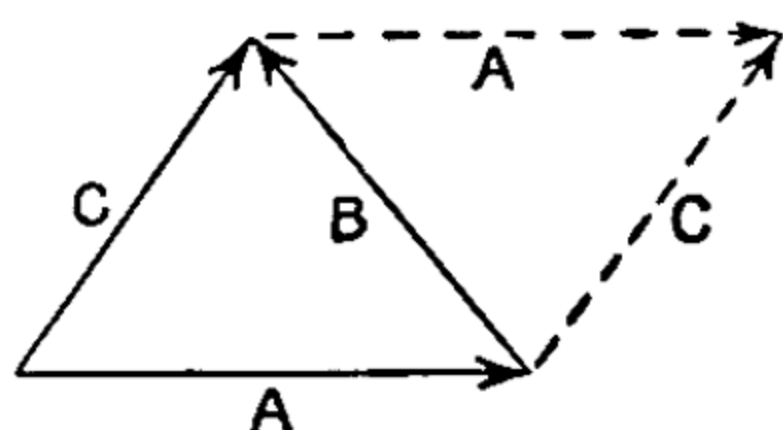


Fig. 3

This geometric sum is *commutative*, like an ordinary sum; i.e. the value is independent of the order in which the summation is performed. As is seen from fig. 3, we obtain the same vector, apart from the immaterial parallel displacement, by laying off **A** from the end point of **B** and drawing a line from the initial point of **B** to the terminus of **A**. Further, a vector sum is *associative*; i.e. in the case of a sum of several vectors, the individual terms may be grouped in any arbitrary manner:

$$\begin{aligned} \mathbf{S} &= (\mathbf{A} + \mathbf{B}) + \mathbf{C} = \mathbf{A} + (\mathbf{B} + \mathbf{C}) \\ &= \mathbf{B} + (\mathbf{A} + \mathbf{C}) \quad . . . (2) \end{aligned}$$

The proof of this is evident from fig. 4.

The sum of two vectors having the same direction and sense is obtained by addition of their lengths, the direction remaining the same. From this follows at once the definition of the product of a vector and a pure number. According to the meaning of multiplication,  $m\mathbf{A}$  means the sum of  $m$  terms, each one being **A**, i.e. a vector having the same direction, but  $m$  times as long. This leads to an important representation for a vector: if the direction is specified by a vector of length 1, any vector having this direction may be represented by multiplying this unit vector  $\mathbf{e}$  by the magnitude of **A**, which may be designated by the corresponding Roman letter, or by  $|\mathbf{A}|$ :

$$\mathbf{A} = A\mathbf{e} = |\mathbf{A}|\mathbf{e}. \quad . . . . . (3)$$

If we think of a vector as being compounded from three vectors having

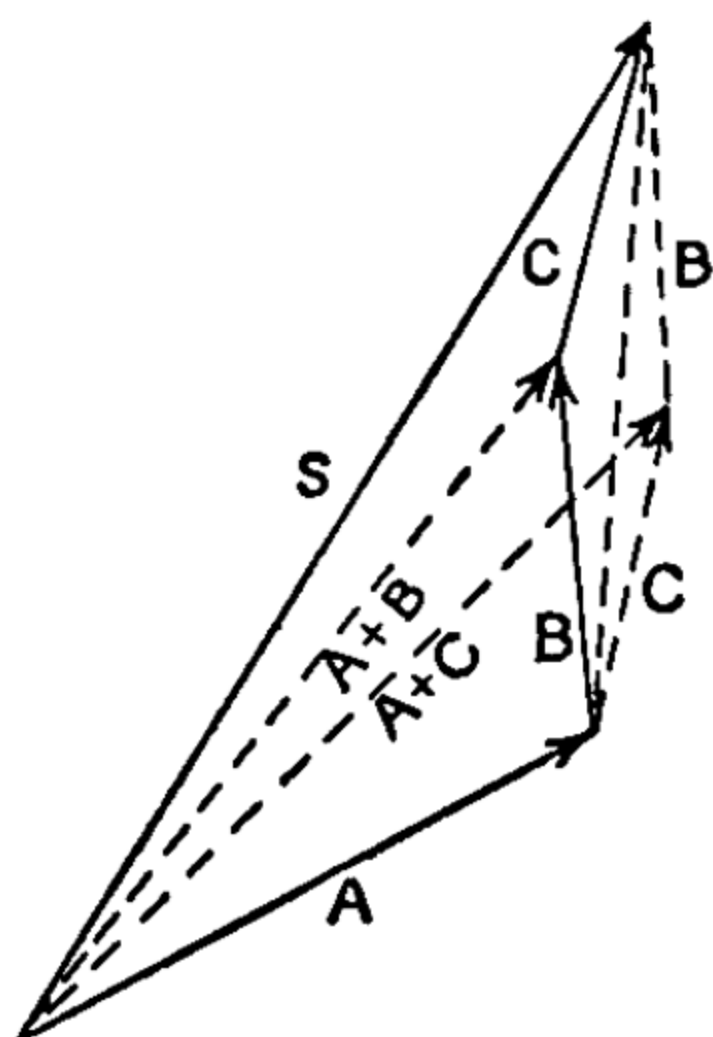


Fig. 4



the directions of the axes of a rectangular co-ordinate system, and if we denote the lengths of these vectors (which we shall call the rectangular components of  $\mathbf{A}$ ), by  $A_x, A_y, A_z$ , and if  $i, j, k$  are the unit vectors in the directions of the axes, then the original vector is given, in terms of its components, as

$$\mathbf{A} = A_x i + A_y j + A_z k. \quad . \quad . \quad . \quad . \quad (4)$$

The *difference* of the vectors  $\mathbf{C}$  and  $\mathbf{A}$  is a vector  $\mathbf{B}$  which when added to  $\mathbf{A}$  gives the vector  $\mathbf{C}$ . As is seen from fig. 3 (p. 9), we obtain  $\mathbf{B}$  by laying off  $\mathbf{A}$ , with its direction reversed, from the terminus of  $\mathbf{C}$ , and then connecting the initial point of  $\mathbf{C}$  with the end point of  $\mathbf{A}$ . From this it follows at once that the meaning of the multiplication of a vector by  $(-1)$  is merely *reversal of direction*, for, retaining the formal laws of ordinary arithmetic, we may look upon the subtraction of the vector  $\mathbf{A}$  as the addition of a vector  $-\mathbf{A}$ .

*Ex. 1.* Express, by means of an equation, the fact that three vectors  $\mathbf{A}$ ,  $\mathbf{B}$ , and  $\mathbf{C}$  form a closed triangle, all three vectors describing the perimeter in the same direction.

*Ex. 2.* Express each of the following by an equation:

(a) Two vectors  $\mathbf{A}$  and  $\mathbf{B}$  are parallel.

(b) Three vectors  $\mathbf{A}$ ,  $\mathbf{B}$  and  $\mathbf{C}$  are coplanar.

### 3. The Scalar Product of two Vectors.

In physics there arise certain combinations of vectors which possess the more important properties of products. We define as the *scalar product of two vectors*, a number (scalar) equal to the product of the magnitudes of the two vectors multiplied by the cosine of the angle included between them. The scalar product is positive if the included angle is acute, and negative if the angle is obtuse. Then, if we agree to indicate the scalar product by writing the two vectors alongside each other\* the definition of the scalar product reads

$$\mathbf{AB} = AB \cos(\mathbf{AB}), \quad . \quad . \quad . \quad . \quad . \quad (5)$$

where  $(\mathbf{AB})$  is the angle included between the two vectors. Since  $A \cos(\mathbf{AB})$  is equal to the projection of the vector  $\mathbf{A}$  upon the direction of  $\mathbf{B}$ , the scalar product may also be defined as the product of the magnitude of one vector by the projection of the other upon it. A simple expression follows for this projection of a vector  $\mathbf{A}$  upon a direction given by the unit vector  $\mathbf{e}$ : it is merely the scalar product  $\mathbf{Ae}$ .

The scalar product has the property of commutativity:

$$\mathbf{AB} = \mathbf{BA}, \quad . \quad . \quad . \quad . \quad . \quad (6)$$

\* It is sometimes convenient to write the scalar product within brackets, in which case the brackets used are round; e.g. the product of the two scalar products  $\mathbf{AB}$  and  $\mathbf{CD}$  may be written  $(\mathbf{AB})(\mathbf{CD})$ . The scalar product itself is written  $(\mathbf{AB})$  or  $(\mathbf{A} \cdot \mathbf{B})$  by some authors.



for, in the defining equation (5) (p. 10), the order of the factors is immaterial.

The scalar product possesses, further, the most important property of an ordinary number product—that of being distributive with respect to addition. This is expressed by the equation

$$\mathbf{A}(\mathbf{B} + \mathbf{C} + \mathbf{D} + \dots) = \mathbf{AB} + \mathbf{AC} + \mathbf{AD} + \dots \quad (7)$$

It may be seen from fig. 5, that the projection, on the direction of  $\mathbf{A}$ , of the sum of the three vectors is the same as the sum of the projections of the vectors, whence, on account of the way the scalar product was defined, equation (7) is verified.

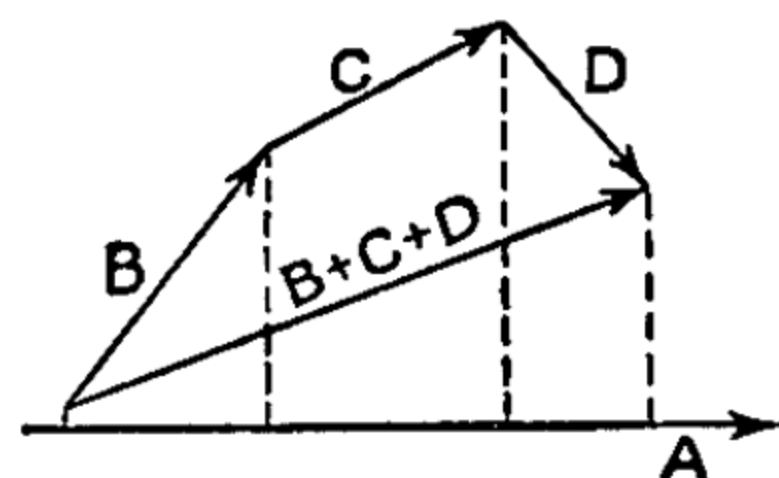


Fig. 5

However, the scalar product of two vectors differs from the product of two numbers in one respect. The latter vanishes only if one or both of the factors are zero; but the scalar product vanishes also when  $\cos(\mathbf{AB})$  is zero, i.e. if the two vectors are perpendicular. Thus, for the three unit vectors of a rectangular co-ordinate system, denoted as before by  $i, j, k$ , we have

$$ij = jk = ki = 0. \quad (8)$$

The scalar product of a given vector by itself is termed the square of the vector. Since  $\cos(\mathbf{AB}) = 1$  in this case, the result is simply the square of the magnitude of the vector:

$$\mathbf{A}^2 = \mathbf{AA} = A^2, \quad (9)$$

and, reciprocally, the magnitude may be written

$$A = |\sqrt{\mathbf{A}^2}|. \quad (10)$$

Thus, for the unit vectors,

$$ii = jj = kk = 1. \quad (11)$$

If the rectangular components of two vectors are given,

$$\mathbf{A} = A_x i + A_y j + A_z k \quad \text{and} \quad \mathbf{B} = B_x i + B_y j + B_z k,$$

we obtain, on account of the distributive property, by multiplying out in the usual manner,

$$\mathbf{AB} = A_x B_x ii + A_y B_y ij + \dots,$$

which, by equations (8) and (11), reduces to

$$\mathbf{AB} = A_x B_x + A_y B_y + A_z B_z. \quad (12)$$

Thus the scalar product of two vectors is equal to the sum of the products of the corresponding components.

In rectangular co-ordinates, the components are identical with the projections upon the axes, so that

$$A_x = i \mathbf{A}. \quad . . . . . (13)$$

It is immediately evident from this result that if  $\mathbf{C}$  is the vector sum of  $\mathbf{A}$  and  $\mathbf{B}$ , then each component of  $\mathbf{C}$  is equal to the sum of the corresponding components; for if equation (1) be multiplied by  $i$  there results immediately

$$C_x = A_x + B_x. \quad . . . . . (14)$$

Equation (13) is no longer true for oblique co-ordinates, as may be seen from fig. 6, for the two-dimensional case. If we regard as com-

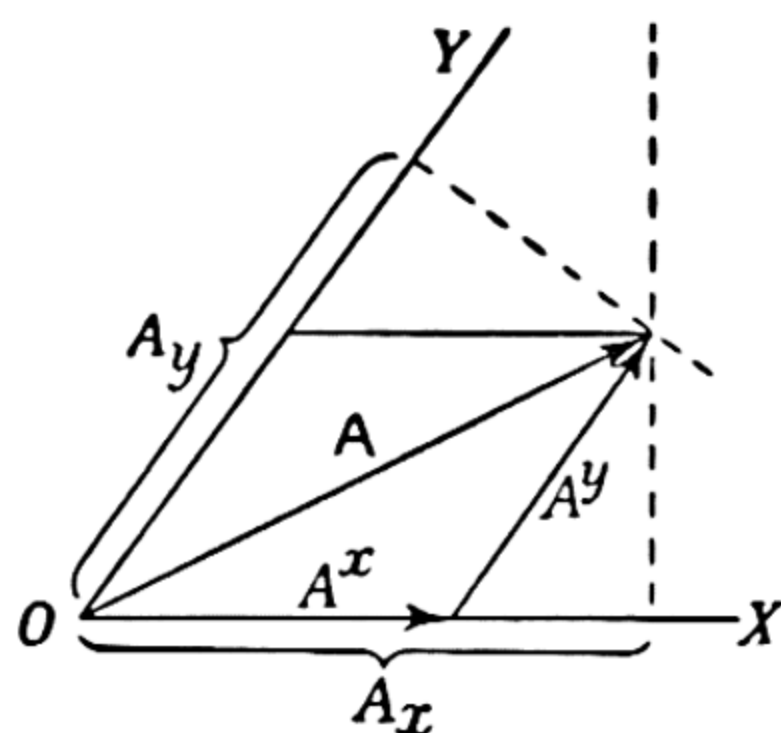


Fig. 6

ponents of  $\mathbf{A}$  the partial vectors in the directions of the axes (from which  $\mathbf{A}$  may be obtained by vector addition), then it is evident that these partial vectors are *not* equal to the projections upon the axes. But the terminus of  $\mathbf{A}$  may be determined uniquely by specifying the projections upon the axes, as well as by giving the partial vectors. The former are, therefore, in a sense, also components of  $\mathbf{A}$ . Because of their different behaviour with respect to linear transformations, the components of the first kind, which

yield  $\mathbf{A}$  by vector addition, are called the *contravariant* components, and the projections are called the *covariant* components of the vector. This distinction is important in cases where the use of oblique co-ordinates is essential, as in the Generalized Theory of Relativity.

*Ex. 3.* Give the geometric significance of

$$(\mathbf{A} + \mathbf{B})^2 = \mathbf{A}^2 + 2\mathbf{A}\mathbf{B} + \mathbf{B}^2.$$

*Ex. 4.* What is the meaning of  $(\mathbf{A} + \mathbf{B})(\mathbf{A} - \mathbf{B})$  for the case where  $\mathbf{A}^2 = \mathbf{B}^2$ ?

*Ex. 5.* Calculate the angle between the two vectors

$$\mathbf{s}_n = \cos \alpha_n i + \cos \beta_n j + \cos \gamma_n k \quad (n = 1, 2).$$

#### 4. The Vector Product of two Vectors. The Directed Plane Area as a Vector.

Besides the scalar product, there is another equally important product-like combination of two vectors, which is itself a vector. The *vector product* of two vectors  $\mathbf{A}$  and  $\mathbf{B}$  is defined as a vector  $\mathbf{P}$  which is perpendicular to the plane determined by  $\mathbf{A}$  and  $\mathbf{B}$ , and whose magnitude is equal to the area of the parallelogram formed by  $\mathbf{A}$  and  $\mathbf{B}$ , i.e. equal to  $AB \sin(\angle \mathbf{A}\mathbf{B})$ . The sign (sense) of  $\mathbf{P}$  is so determined that, sighting along  $\mathbf{P}$ , the shortest rotation from  $\mathbf{A}$  (the first factor), to-

ward **B** (the second factor), is clockwise (fig. 7). This stipulation denies the vector product the property of commutativity possessed by the scalar product; for if we reverse the order of the factors, then the turning of **B** toward **A** must be clockwise when looking in the direction of **P**, which means that **P** is now reversed. The vector product is indicated by writing the two vectors alongside each other, and enclosing them in square brackets.\* Thus

$$\mathbf{P} = [\mathbf{AB}] = -[\mathbf{BA}]. \quad . \quad . \quad (15)$$

The magnitude of the product vector is

$$P = AB \sin(\angle \mathbf{AB}). \quad . \quad . \quad (15')$$

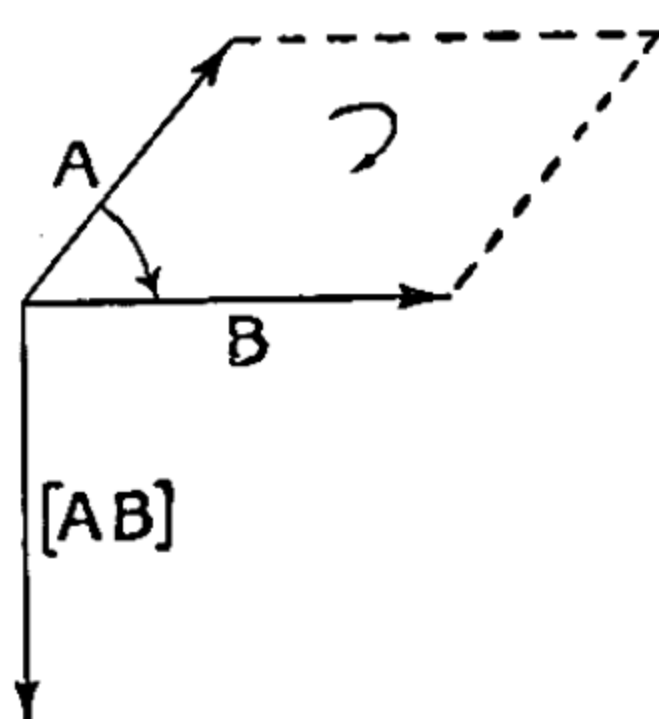


Fig. 7

The most important property of the vector product, without which the use of the term 'product' would not be justified, is that of distributivity with respect to addition. That is,

$$[\mathbf{A}(\mathbf{B} + \mathbf{C} + \mathbf{D} + \dots)] = [\mathbf{AB}] + [\mathbf{AC}] + [\mathbf{AD}] + \dots \quad (16)$$

The proof of this formula is somewhat cumbersome, but not difficult, and will not be given here.

Just as the scalar product of two non-vanishing vectors becomes zero when the two vectors are perpendicular, so the vector product vanishes if the factors are parallel. The equation  $[\mathbf{AB}] = 0$  expresses the fact that **A** and **B** are parallel, provided that **A** and **B** differ from zero. In the same way,  $[\mathbf{AA}] = 0$ . For the unit vectors along the axes,

$$[i i] = [j j] = [k k] = 0. \quad . \quad . \quad . \quad (17)$$

$$[i j] = k; [j k] = i; [k i] = j. \quad . \quad . \quad . \quad (18)$$

Using these relationships, the application of the distributive property gives the following representation of the vector product in terms of the components:

$$\begin{aligned} [\mathbf{AB}] &= [(A_x i + A_y j + A_z k)(B_x i + B_y j + B_z k)] \\ &= (A_y B_z - A_z B_y) i + (A_z B_x - A_x B_z) j + (A_x B_y - A_y B_x) k. \end{aligned}$$

This may be written as a determinant:

$$[\mathbf{AB}] = \begin{vmatrix} i & j & k \\ A_x & A_y & A_z \\ B_x & B_y & B_z \end{vmatrix} \quad . \quad . \quad . \quad (19)$$

The change in sign of the vector product with reversal of the order of the factors manifests itself here as the well-known law that the interchange of two rows changes the sign of the determinant.

\* Besides the square brackets, the notation  $\mathbf{A} \times \mathbf{B}$  is often used.



The specification of the product vector  $\mathbf{P}$  gives the following information concerning the parallelogram formed by the two vectors  $\mathbf{A}$  and  $\mathbf{B}$ : (1) its position in space (the plane perpendicular to  $\mathbf{P}$ ); (2) the relative position of the sides  $\mathbf{A}$  and  $\mathbf{B}$  with respect to one another; (3) the area, numerically equal to the length of  $\mathbf{P}$ . The form of the parallelogram is not given by  $\mathbf{P}$ . In looking upon the form as being unimportant, we are recognizing the fact that every portion of a plane may be represented by a vector whose direction and magnitude are uniquely determined, according to the above rules, by the orientation, direction of description, and area of the figure.

For the vectors  $\mathbf{S}_i$  representing the faces of a closed polyhedron, the following theorem holds: If the representative vector  $\mathbf{S}_i$  is assigned to each face in such manner that it is directed towards the outside of the solid in every case, then the sum of these vectors vanishes:

$$\Sigma \mathbf{S}_i = 0 \quad \text{for closed polyhedra.} \quad . . . (20)$$

This theorem is easily verified for a tetrahedron. If we denote the

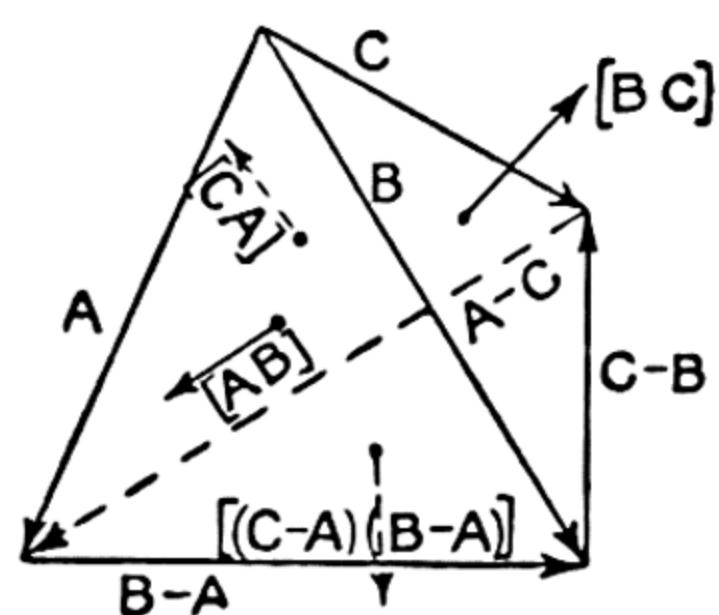


Fig. 8

three vectors emanating from the vertex (fig. 8) by  $\mathbf{A}$ ,  $\mathbf{B}$  and  $\mathbf{C}$ , then the three faces passing through this vertex are  $[\mathbf{AB}]$ ,  $[\mathbf{BC}]$  and  $[\mathbf{CA}]$ . The base is represented by  $[(\mathbf{C} - \mathbf{A})(\mathbf{B} - \mathbf{A})]$ . If this is expanded, and the resulting vector products added to the above three, the result is zero. Now any polyhedron may be subdivided into a number of tetrahedra. The above theorem holds for each of these component tetrahedra, and every surface in-

serted by the subdivision in the interior of the solid enters into the calculation twice, with oppositely directed normals, so that in forming the sum, the extra surfaces formed by cutting the original solid into tetrahedra cancel out, leaving only the sum over the external surfaces. We can go a step further: if we have an arbitrary closed curved surface, we can approximate to it by a polyhedron whose faces are portions of the tangent planes. In the limit, when the faces  $d\mathbf{S}$  become vanishingly small, the polyhedron and the actual surface coincide, and the sum passes into an integral

$$\oint d\mathbf{S} = 0 \quad \text{for closed surfaces.} \quad . . . (21)$$

The sign  $\oint$  indicates integration over a closed region. A closed surface of this kind is often termed a "shell"; this type of integral, which may be termed a "shell integral", will be met with often in vector analysis.

In spite of the fact that we can assign the product vector  $\mathbf{P}$  uniquely to the parallelogram formed by two vectors, there is a certain difference between the parallelogram and its representative vector. If, namely, the parallelogram be reflected in its own plane, the direction of one side with respect to the other remains unaltered, but the sign of the product vector changes. This fundamental difference between vectors like displacements (polar vectors) and those which represent a direction of rotation (axial vectors) comes out clearly when we change from a right-handed to a left-handed system of co-ordinates. But since such a change is generally unnecessary and superfluous, this difference is of no consequence in actual calculation. On the other hand, in the four-dimensional vector analysis of the Theory of Relativity, the distinction is of essential importance; in that case a polar vector has four components, while an axial vector has six.

*Ex. 6.* Verify (16), using an expression containing only two terms.

*Ex. 7.* What is the value of  $[\mathbf{AB}]^2 + (\mathbf{AB})^2$ ?

*Ex. 8.* In solid analytical geometry, all theorems concerning lines and planes may be derived with little calculation if we operate with the radius vector  $\mathbf{r}$  and specify directions by unit vectors, instead of using Cartesian co-ordinates. In this manner give:

(a) the equation of a plane whose normal direction is that of  $\mathbf{n}$  and whose distance from the origin is  $p$ ,

(b) the distance of a point  $\mathbf{r}_0$  from this plane,

(c) the equation of the plane passing through three points  $\mathbf{r}_1, \mathbf{r}_2, \mathbf{r}_3$ .

## 5. Multiple Products.

(a) *Product of a vector with the scalar product of two other vectors*

Since a scalar product is an ordinary number, the product of a vector  $\mathbf{A}$  with the scalar product  $\mathbf{BC}$  means the multiplication of  $\mathbf{A}$  with this number, according to the rule given in § 2 (p. 9), i.e. a vector having the direction of  $\mathbf{A}$ . This is written  $\mathbf{A}(\mathbf{BC})$  or  $(\mathbf{BC})\mathbf{A}$ . It is to be noticed that  $(\mathbf{AB})\mathbf{C}$  has an entirely different meaning from  $\mathbf{A}(\mathbf{BC})$ , viz. the former is a vector in the direction of  $\mathbf{C}$ .

(b) *Scalar product of a vector with the vector product of two other vectors*

If  $\mathbf{P} = [\mathbf{AB}]$ , then the product  $\mathbf{CP} = \mathbf{C}[\mathbf{AB}]$  is the volume of the parallelepiped having the three vectors  $\mathbf{A}$ ,  $\mathbf{B}$ , and  $\mathbf{C}$  as contiguous edges, for the magnitude of  $\mathbf{P}$  is equal to the area of the parallelogram formed by  $\mathbf{A}$  and  $\mathbf{B}$  (fig. 9). Thus

$$\mathbf{CP} = CP \cos(\mathbf{CP}),$$

is the volume of the parallelepiped, since  $C \cos(\mathbf{CP})$  is equal to its altitude. One might equally well consider the base of the parallelepiped



to be formed by the vectors  $\mathbf{C}$  and  $\mathbf{A}$ , and then form the scalar product of this quantity with  $\mathbf{B}$ . This means that the volume can also be

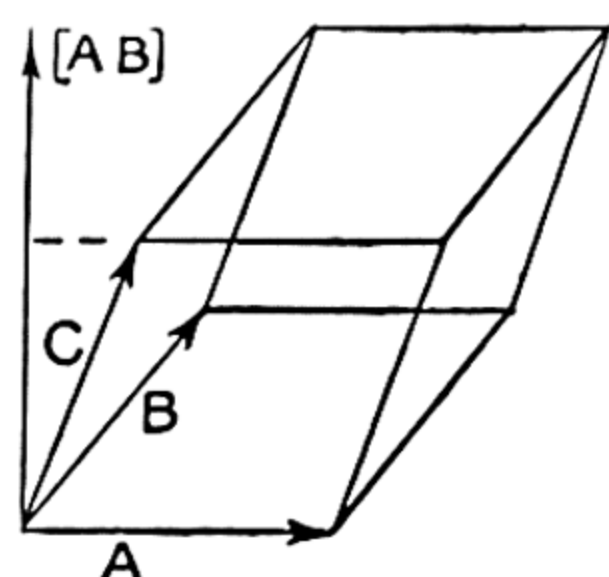


Fig. 9

represented by the product  $[\mathbf{CA}]\mathbf{B}$ . Because of this property, the square brackets are unnecessary, and from now on such a triple product will be denoted simply by writing the three factors alongside each other. However, the sign of the *vector product* changes each time the cyclic order of the factors is changed. The effect of interchanging two factors is seen from fig. 9.  $[\mathbf{BA}]$  is a vector directed downward, and hence  $\cos(\mathbf{CP})$  becomes negative. We may

set up the following rule for the sign of the triple scalar product: The product  $\mathbf{ABC}$  is positive if the three vectors are relatively arranged like the axes of a right-handed co-ordinate system. Thus

$$\mathbf{ABC} = \mathbf{BCA} = \mathbf{CAB} = -\mathbf{ACB} = -\mathbf{BAC} = -\mathbf{CBA}. \quad (22)$$

If the components of the three vectors are given, then

$$\mathbf{ABC} = (A_x \mathbf{i} + A_y \mathbf{j} + A_z \mathbf{k}) \{ (B_y C_z - B_z C_y) \mathbf{i} + (B_z C_x - B_x C_z) \mathbf{j} + (B_x C_y - B_y C_x) \mathbf{k} \}.$$

This may be written as a determinant:

$$\mathbf{ABC} = \begin{vmatrix} A_x & A_y & A_z \\ B_x & B_y & B_z \\ C_x & C_y & C_z \end{vmatrix} \cdot \cdot \cdot \cdot \cdot \quad (23)$$

Equations (22) thus express the well-known property of determinants according to which the interchange of two rows causes a change in algebraic sign.

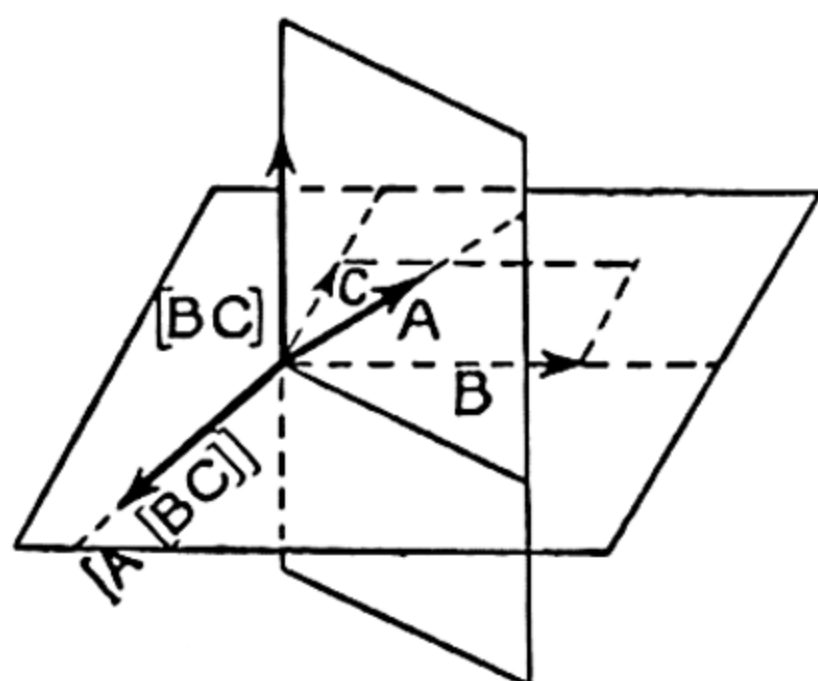


Fig. 10

(c) *Vector product of a vector with the vector product of two other vectors*

If  $\mathbf{P} = [\mathbf{BC}]$ , then the vector product

$$\mathbf{R} = [\mathbf{AP}] = [\mathbf{A}[\mathbf{BC}]]$$

signifies a vector lying in the plane of  $\mathbf{B}$  and  $\mathbf{C}$ , for  $\mathbf{P} = [\mathbf{BC}]$  is perpendicular to this plane, but  $\mathbf{R} = [\mathbf{AP}]$  is in turn perpendicular to  $\mathbf{P}$ , and so falls in the plane of  $\mathbf{B}$  and  $\mathbf{C}$  (fig. 10).

Since any vector  $\mathbf{V}$  lying in the plane of  $\mathbf{B}$  and  $\mathbf{C}$  may be written in the form

$$\mathbf{V} = \alpha \mathbf{B} + \beta \mathbf{C}$$

by suitably choosing the scalar multipliers  $\alpha$  and  $\beta$ ,  $\mathbf{R}$  must have this form, and it remains to determine the numbers  $\alpha$  and  $\beta$ . To this end, we introduce a co-ordinate system in such manner that the  $x$ -axis is in the direction of  $\mathbf{C}$  and the  $y$ -axis is in the plane of  $\mathbf{B}$  and  $\mathbf{C}$ . This may be done without restricting the generality of the discussion. We then have

$$\begin{aligned}\mathbf{A} &= A_x \mathbf{i} + A_y \mathbf{j} + A_z \mathbf{k}, \\ \mathbf{B} &= B_x \mathbf{i} + B_y \mathbf{j}, \\ \mathbf{C} &= C_x \mathbf{i}.\end{aligned}$$

The vector product  $[\mathbf{BC}]$  then assumes the simple form

$$[\mathbf{BC}] = -B_y C_x \mathbf{k},$$

and 
$$[\mathbf{A} [\mathbf{BC}]] = -A_y B_y C_x \mathbf{i} + A_x B_y C_x \mathbf{j}.$$

If the vector  $A_x B_x C_x \mathbf{i}$  be added and subtracted, there results

$$[\mathbf{A} [\mathbf{BC}]] = A_x C_x (B_x \mathbf{i} + B_y \mathbf{j}) - (A_x B_x + A_y B_y) C_x \mathbf{i}.$$

This is equivalent to the important formula

$$[\mathbf{A} [\mathbf{BC}]] = \mathbf{B}(\mathbf{AC}) - \mathbf{C}(\mathbf{AB}). \quad . \quad . \quad . \quad (24)$$

The coefficients  $\alpha$  and  $\beta$  are thus the scalar products  $\mathbf{AC}$  and  $-\mathbf{AB}$ .

With the aid of this formula, complex products may be converted to simpler forms. Take, for example, the scalar product of two vector products:

$$P = [\mathbf{AB}] [\mathbf{CD}].$$

We put  $[\mathbf{AB}] = \mathbf{E}$  and obtain, by (22) (p. 16)

$$P = \mathbf{E} [\mathbf{CD}] = -\mathbf{C} [\mathbf{ED}] = -\mathbf{C} [[\mathbf{AB}] \mathbf{D}],$$

and then, using (24),

$$P = \mathbf{C} \{ \mathbf{A}(\mathbf{DB}) - \mathbf{B}(\mathbf{AD}) \}.$$

Hence 
$$[\mathbf{AB}] [\mathbf{CD}] = (\mathbf{AC}) (\mathbf{BD}) - (\mathbf{BC}) (\mathbf{AD}). \quad . \quad . \quad (25)$$

*Ex. 9.* Transform  $[[\mathbf{AB}][\mathbf{CD}]]$ .

## 6. Differentiation of a Vector with respect to a Scalar; Application to the Theory of Space Curves.

Let a vector  $\mathbf{v}$  be a continuous function of a continuous scalar variable  $u$ :

$$\mathbf{v} = \mathbf{v}(u).$$

If the variable  $u$  is increased by  $\Delta u$ , the vector will change by an amount

$$\Delta \mathbf{v} = \mathbf{v}(u + \Delta u) - \mathbf{v}(u),$$

and, in complete analogy with scalar functions, we define the derivative  $d\mathbf{v}/du$  as the limit

$$\mathbf{v}' = \frac{d\mathbf{v}}{du} = \lim_{\Delta u \rightarrow 0} \frac{\mathbf{v}(u + \Delta u) - \mathbf{v}(u)}{\Delta u}. \quad (26)$$

Derivatives of a higher order are defined similarly, e.g.:

$$\mathbf{v}'' = \frac{d\left(\frac{d\mathbf{v}}{du}\right)}{du} = \frac{d^2\mathbf{v}}{du^2} = \lim_{\Delta u \rightarrow 0} \frac{\mathbf{v}'(u + \Delta u) - \mathbf{v}'(u)}{\Delta u}. \quad (27)$$

If the position vector  $\mathbf{r}$  of the points on a space curve be given as a function of the length of arc  $s$ , measured from a given initial point, then the magnitude of  $\Delta\mathbf{r}$  is identical with  $\Delta s$ , and the limit of  $\Delta\mathbf{r}/\Delta s$  is a vector of length 1, having the direction of the tangent to the curve (fig. 11). This *unit tangent vector* is denoted by  $\mathbf{t}$ , and we have

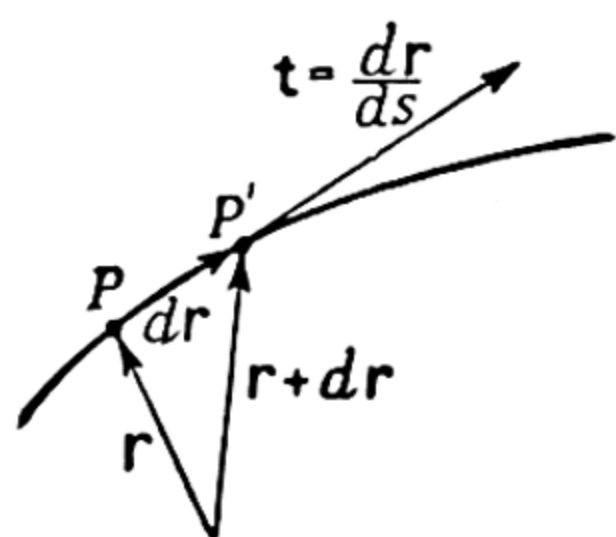


Fig. 11

$$\mathbf{t} = \frac{d\mathbf{r}}{ds}. \quad (28)$$

The rules for the differentiation of products also correspond to those for ordinary scalar functions, as may be seen readily by writing the derivatives as limits. In the case of the vector product, however, the order of the factors is important. We thus have

$$\frac{d(\mathbf{AB})}{du} = \mathbf{A} \frac{d\mathbf{B}}{du} + \frac{d\mathbf{A}}{du} \mathbf{B} = \mathbf{A} \frac{d\mathbf{B}}{du} + \mathbf{B} \frac{d\mathbf{A}}{du}, \quad (29)$$

but 
$$\frac{d[\mathbf{AB}]}{du} = \left[ \mathbf{A} \frac{d\mathbf{B}}{du} \right] + \left[ \frac{d\mathbf{A}}{du} \mathbf{B} \right] = \left[ \mathbf{A} \frac{d\mathbf{B}}{du} \right] - \left[ \mathbf{B} \frac{d\mathbf{A}}{du} \right]. \quad (30)$$

If a vector does not change in length, then, since  $\mathbf{v}^2 = \text{const.}$ ,

$$\frac{d\mathbf{v}^2}{du} = 2\mathbf{v} \cdot \frac{d\mathbf{v}}{du} = 0.$$

Since neither  $d\mathbf{v}/du$  nor  $\mathbf{v}$  is to vanish, this means that the derivative of a vector of constant length is perpendicular to the vector. This is also evident geometrically, for if the length is constant, the end point of the vector is restricted to move on a sphere. If the increment is infinitesimal, it is tangent to the sphere, and hence is perpendicular to the vector itself.

This result may be applied to the unit tangent  $\mathbf{t}$ . Since this vector is always of unit length, its derivative must always be



perpendicular to  $\mathbf{t}$ , and so must be a vector in the normal plane to the curve. But this derivative, being the vector difference of two consecutive tangent vectors, must lie in the *osculating plane* formed by the latter, and so its direction is that of the *principal normal*, which direction we designate by the unit vector  $\mathbf{n}$ . In order to calculate the magnitude of the vector  $d\mathbf{t}/ds$ , we note that the curve, in the neighbourhood of two consecutive tangents, corresponding to three neighbouring points, may be replaced by the circle of curvature, whose centre  $M$  is determined by the intersection of the perpendiculars to the two consecutive tangents (fig. 12). The angle  $d\phi$  between these tangents is the same as that between the two radii of the circle of curvature. If  $\rho$  is the radius of this circle then, in the limit,

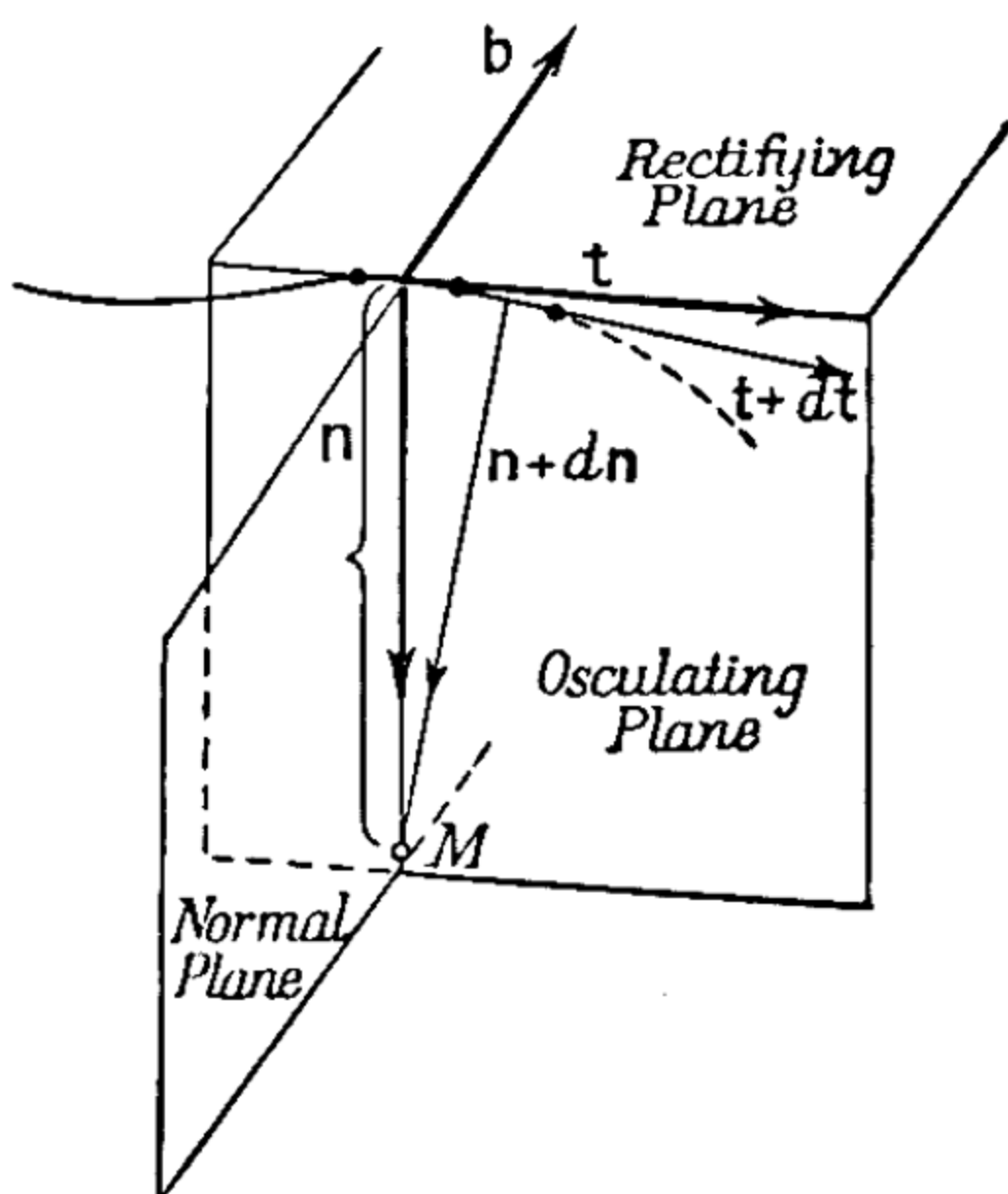


Fig. 12

$$ds = \rho d\phi. \quad \dots \dots \dots (31)$$

On the other hand,  $|d\mathbf{t}| = d\phi$ , whence

$$\left| \frac{d\mathbf{t}}{ds} \right| = \frac{1}{\rho}.$$

Thus is obtained the important equation

$$\frac{d\mathbf{t}}{ds} = \frac{d^2\mathbf{r}}{ds^2} = \frac{\mathbf{n}}{\rho} \quad \dots \dots \dots (32)$$

The value of the *curvature* is obtained from this result by squaring (scalar product) and extracting the root:

$$k = \frac{1}{\rho} = \sqrt{\left(\frac{d^2\mathbf{r}}{ds^2}\right)^2} = \sqrt{\left(\frac{d^2x}{ds^2}\right)^2 + \left(\frac{d^2y}{ds^2}\right)^2 + \left(\frac{d^2z}{ds^2}\right)^2}. \quad (33)$$

The unit vector  $\mathbf{b}$  perpendicular to the osculating plane, gives the direction of the *binormal*. If we stipulate that the three vectors  $\mathbf{t}$ ,  $\mathbf{n}$  and  $\mathbf{b}$ , in this order, form a right-handed triad, then

$$\mathbf{b} = [\mathbf{t}\mathbf{n}]. \quad \dots \dots \dots (34)$$

For a plane curve, the osculating plane (and therefore the binormal) have constant direction, so that  $d\mathbf{b}/ds$  vanishes. For a non-plane (skew) curve, this





which is called the gradient of  $u$ , and is written  $\text{grad } u$ :

$$\text{grad } u = \frac{\partial u}{\partial x} i + \frac{\partial u}{\partial y} j + \frac{\partial u}{\partial z} k, \quad . . . . (36)$$

that is,

$$du = ds(\text{grad } u). \quad . . . . (37)$$

The meaning of the gradient may be seen in this way: Leaving discontinuous functions out of consideration, and dealing only with continuous point-functions, we can connect all points in the region having the same value of  $u$  by means of surfaces—so-called *level surfaces* (fig. 13). Each of these surfaces is characterized by the fact that displacements wholly within the surface do not alter the value of  $u$ . If, then, we let  $ds_0$  lie in one of the level surfaces, we have

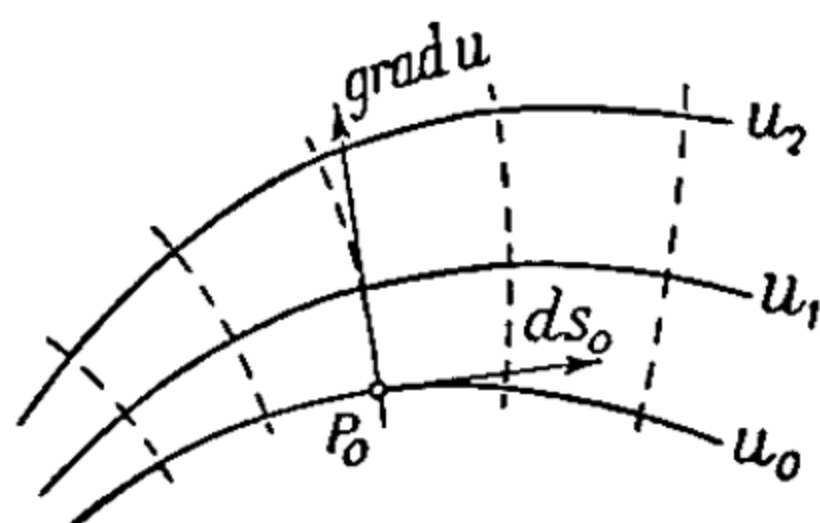


Fig. 13

If, then, we let  $ds_0$  lie in one of the level surfaces, we have

$$ds_0(\text{grad } u) = 0.$$

But since neither  $ds_0$  nor  $\text{grad } u$  is to vanish, this means that the vector  $\text{grad } u$  is perpendicular to the level surface.

Further, from the definition of the scalar product, it follows that the increment  $du$  has its greatest value—the magnitude of  $ds$  remaining constant—when  $ds$  and  $\text{grad } u$  are in the same direction, or, in other words, the vector  $\text{grad } u$  gives the direction of greatest change (“slope”) of the point-function  $u$ —hence the name “gradient”. According to (37), its magnitude is equal to  $du/ds$  if  $ds$  is perpendicular to the level surface. The change of  $u$  in any other direction is obtained, according to (37), by multiplying by the cosine of the angle  $\phi$  between the gradient and the direction

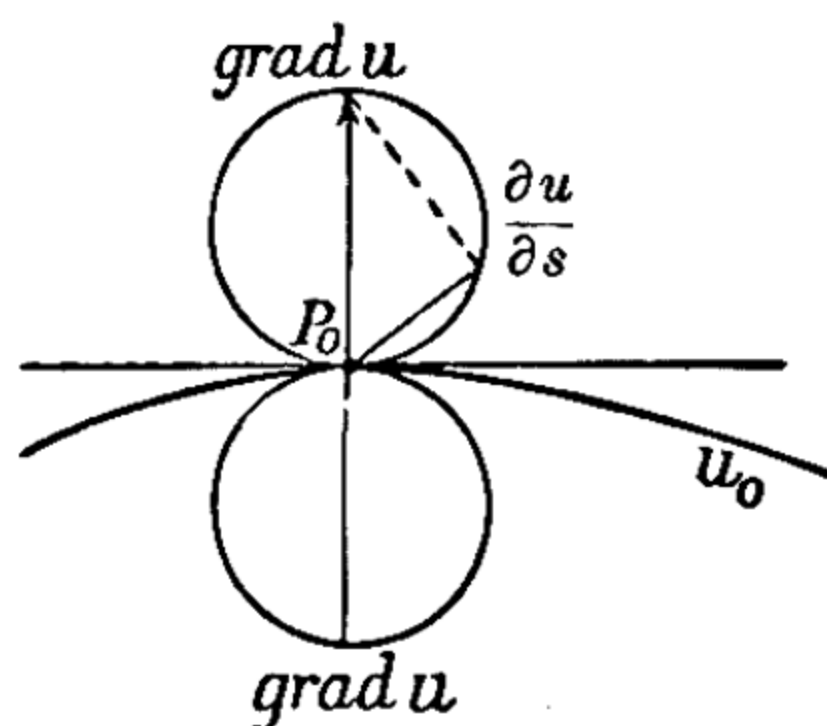


Fig. 14

of displacement. This suggests a simple graphical construction for the several values of  $du$ , for a  $ds$  of fixed magnitude (fig. 14). Two lines of length equal to the magnitude of  $\text{grad } u$  are drawn through  $P_0$  and normal to the level surface, one above and one below, and the spheres having these two lines as diameters are described. Then a secant drawn in any direction gives the magnitude of the change in that direction, in relation to the change  $du$  in the direction of the gradient. It must be remembered that  $ds$  is held constant here.

The derivative of  $u$  in the direction of  $s$  is denoted by  $\partial u / \partial s$ , and is thus obtained by projection of  $\text{grad } u$  on the direction  $s$ .

Numerous examples of the physical occurrence of the gradient concept will be met with in many sections of this work. One example may be given now. A weather map has marked upon it the lines of equal air pressure (isobars). The direction of the wind is then given, apart from the effect of the earth's rotation, by the direction of greatest pressure drop, which is perpendicular to the isobars; and the strength of the wind is determined by the magnitude of the pressure drop.

By introducing the gradient we have obtained a *vector field*, that of  $\text{grad } u$ , from a *scalar field*, for  $\text{grad } u$  is a vector magnitude which alters in value from place to place in the region. However, not every vector field may be obtained from a scalar field in this manner. If a

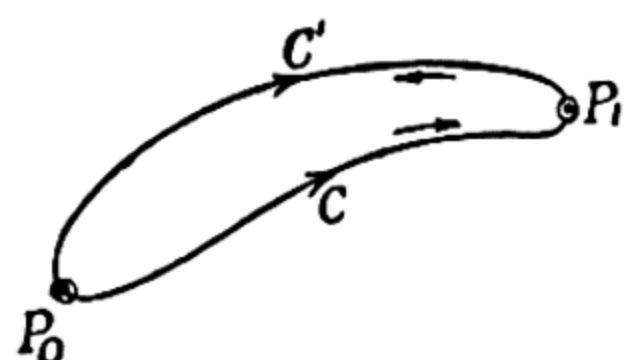


Fig. 15

vector field has this property, a result of great importance follows: Imagine a curve  $C$  drawn in the vector field (fig. 15). Let this curve consist of a succession of line-elements  $ds$ . The *line integral* of a vector  $\mathbf{v}$  along the curve  $C$  is defined as the limit, for  $ds$  infinitely small, of the sum of the scalar products  $\mathbf{v} ds$ , the summation to begin at  $P_0$

and end at  $P_1$ . If now  $\mathbf{v}$  is the gradient of a scalar  $u$ , this line integral between  $P_0$  and  $P_1$  is independent of the form of the curve  $C$ , for the individual magnitudes

$$\mathbf{v} ds = ds (\text{grad } u) = du$$

represent, according to the definition of the gradient, the change in the function  $u$  in moving along a distance  $ds$ . The sum of these changes is the difference between the values of  $u$  at  $P_1$  and at  $P_0$ , and this difference is independent of the path by which one passes from  $P_0$  to  $P_1$ .

If there are two paths,  $C$  and  $C'$ , then reversing the direction of motion along  $C'$  gives with  $C$  a closed curve which is described in one direction. But since the reversal of  $C'$  causes all its line-elements  $ds$  to reverse their sign, and since the value of the line integral is the same on both portions of the curve, the integral along the path vanishes. We thus have the important theorem: *If a vector field is representable as a field of the gradient of a scalar point function, the value of the line integral of the vector, taken between two points in the region, is independent of the path; and the line integral over a closed path vanishes.* Thus

$$\int_{P_0}^{P_1} (\text{grad } u) ds = u_1 - u_0, \quad . \quad . \quad . \quad . \quad (38)$$

and

$$\oint (\text{grad } u) ds = 0. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (38')$$

From the definition (36) of the gradient, a useful rule follows: If  $u$  is a function of a simple point function  $v$ , then, since



$$\frac{\partial u}{\partial x} = \frac{du}{dv} \frac{\partial v}{\partial x},$$

$$\text{grad } u = \frac{du}{dv} \text{grad } v. \quad . \quad . \quad . \quad . \quad (39)$$

*Ex. 12.* What is the meaning of  $\text{grad } r$ ? ( $r$  = distance from origin.)

*Ex. 13.* Let a scalar point function depend only on the distance  $r$  from the origin. Compute  $\text{grad } f(r)$ .

## 8. The Concept of Divergence and Gauss's Theorem.

If an arbitrary vector field is given, we may, as in the analogous case of § 6 (p. 17), wish to determine the change in the field vector  $\mathbf{v}$  accompanying a displacement  $d\mathbf{s}$ . As may readily be seen, the calculation of this change necessitates a knowledge of three gradients, viz. those of the rectangular components, which give the amounts of the changes of the components. We thus arrive at a set of nine scalars (three vectors). Quantities of this nature will not be met with until later, and so we put this question aside, and take up another physically significant differential operation, which leads from a vector field to a scalar field.

In order to understand this operation more clearly, we give the field vector a special meaning. Let the vector  $\mathbf{v}$  represent, in direction and magnitude, the volume of fluid passing, per second, through an area of 1 sq. cm. perpendicular to  $\mathbf{v}$  (fig. 16). The amount of fluid passing through the element of area  $\Delta S$ , each second, is then given by the scalar product  $\mathbf{v} \Delta S$ ; for the quantity passing *through* is given by  $v_n$ , the component of the flow *perpendicular* to the surface element. Let us now enclose a space  $\tau$  in the vector field by means of an arbitrary closed surface. Call the elements of this surface  $dS$  and let them be so oriented that the normals are directed outward. The amount of fluid passing *outward* through an element  $dS$ , in unit time, is therefore  $\mathbf{v} dS$ ; inward flow has the negative sign. Summation of these elementary contributions over the entire surface gives the excess amount of fluid passing to the outside each second. This yields the surface integral  $\oint \mathbf{v} dS$ . If there are no sources within the surface, the integral must vanish. If sources exist within, the integral furnishes a measure of the combined strength of all sources within the bounded volume. The total strength depends also on the volume, and the important quantity is the ratio of the surface integral to the volume. In the case of sources continuously distributed in space, by means of this ratio we can express the source-strength for each element of volume. For the limiting value of the ratio, the expression  $\text{div } \mathbf{v}$  (read "divergence of  $\mathbf{v}$ "), has been



Fig. 16

introduced. Divergence is, then, equivalent to source-strength. Thus

$$\operatorname{div} \mathbf{v} = \lim_{\Delta \tau} \frac{1}{\Delta \tau} \oint \mathbf{v} d\mathbf{S}. \quad . . . . . (40)$$

If the components of  $\mathbf{v}$  are continuous point functions having continuous derivatives, this limit is finite, and independent of the *form* of the volume element. This latter fact is somewhat troublesome to prove, but it is easy to show that the limit is finite. Thus, let the volume element selected be a small sphere of radius  $\rho$  at whose centre the field vector has the value  $\mathbf{v}_0$ . The departures from this value are then, in any case, of the order of magnitude of the radius  $\rho$ . Now the surface element on the sphere is  $\rho^2 d\Omega$ , where  $d\Omega$  is an elementary solid angle. The quantities to be summed in forming the integral are thus of the form  $(\mathbf{v}_0 + \mathbf{w}\rho)d\mathbf{S}$ , where  $\mathbf{w}$  is a variable finite vector, whose further properties need not be given. Since by (21) (p. 14)  $\oint d\mathbf{S} = 0$ , the part of the integral contributed by the first part, viz.  $\oint \mathbf{v}_0 d\mathbf{S} = \mathbf{v}_0 \oint d\mathbf{S}$

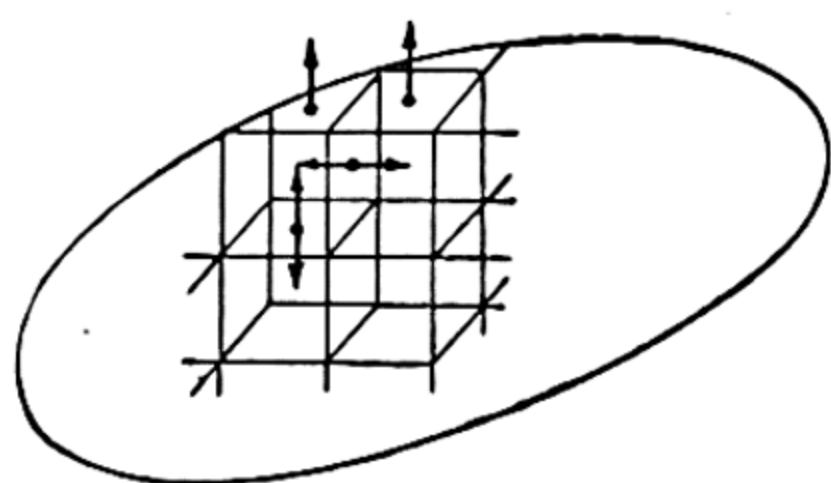


Fig. 17

vanishes, and there remains a contribution from the second part, of the order of  $\rho^3$ . But the volume of the sphere is  $4\pi\rho^3/3$ , so that the result of division by this quantity must be finite.

Let us now consider a finite volume which is bounded by a closed surface (fig. 17). Let this region be sub-

divided into volume elements. The definition of the divergence holds for each of these:

$$\operatorname{div} \mathbf{v} d\tau = \oint \mathbf{v} d\mathbf{S}.$$

Let both sides be summed over all the elements of volume. The left side becomes the volume integral  $\int \operatorname{div} \mathbf{v} d\tau$ . In the right member, all the interior surface elements occur twice (as surfaces of adjacent volume elements) with oppositely directed normals. Their contributions cancel out, since the vector  $\mathbf{v}$ , on account of its continuity, has the same magnitude both times. There remain only the contributions of elements of the outer surface. This yields *Gauss's Theorem*:

$$\oint \mathbf{v} d\mathbf{S} = \int \operatorname{div} \mathbf{v} d\tau. \quad . . . . . (41)$$

That is, *the surface integral of the vector  $\mathbf{v}$ , taken over a closed surface, is equal to the volume integral of the divergence of  $\mathbf{v}$  taken throughout the enclosed volume.* This is sometimes called the Divergence Theorem.

If the hydrodynamical case is taken as a basis for demonstrating



this theorem, the result appears almost self-evident. The theorem, however, is true for any vector field satisfying the conditions given above.

To obtain an expression for the divergence of a vector in terms of its rectangular components, we choose as volume element a rectangular parallelepiped with edges  $dx$ ,  $dy$  and  $dz$ . At  $x$ , an amount of fluid  $dydz(v_x)_x$  passes through the face  $dydz$  (fig. 18). Only the normal component of  $\mathbf{v}$  contributes to the fluid passing through the surface. This fluid enters the volume, since the normal has the direction of the negative  $x$ -axis.

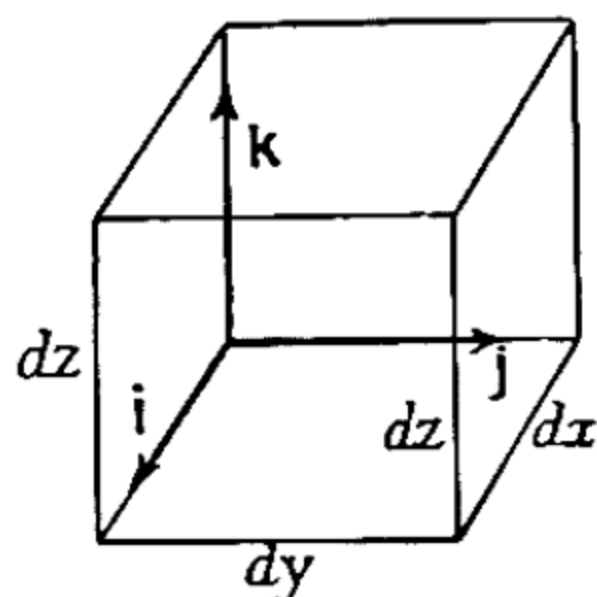


Fig. 18

At  $x + dx$ , an amount  $dydz(v_x)_{x+dx}$  streams out through the opposite face. The normal is here directed along the positive  $x$ -axis. Expanding  $(v_x)_{x+dx}$  by Taylor's Theorem, we have for the contribution of these two surface elements

$$\left( v_x + \frac{\partial v_x}{\partial x} dx - v_x \right) dydz.$$

Analogous contributions are furnished by the two other pairs of opposite faces. Thus we have

$$\oint \mathbf{v} d\mathbf{S} = \left( \frac{\partial v_x}{\partial x} + \frac{\partial v_y}{\partial y} + \frac{\partial v_z}{\partial z} \right) dx dy dz,$$

and, after division by the volume  $dx dy dz$ ,

$$\text{div } \mathbf{v} = \frac{\partial v_x}{\partial x} + \frac{\partial v_y}{\partial y} + \frac{\partial v_z}{\partial z}. \quad \dots \quad (42)$$

*Ex. 14.* Calculate  $\text{div } \mathbf{r}$  from first principles, and also by the formula.

*Ex. 15.* How may the "gold" balance of a country be calculated from the amount of imports and exports? To what extent does this represent an application of Gauss's Theorem? ("Gold" = equivalent of merchandise.)

## 9. The Curl of a Vector, and Stokes's Theorem.

Another equally important differential operation leads from a given vector field not to a scalar field, as in forming the divergence, but to another vector field.

We refer back to the relation (38) (p. 22), according to which the line integral of a vector  $\mathbf{v}$ , taken over a closed curve, vanishes if the vector is the gradient of a scalar field. In general, this integral does not vanish; and its magnitude measures an important property of the field, especially when the integral is taken for infinitesimal sur-



face elements, and referred to unit area. Here, too, it may be shown that the limiting value is independent of the particular form of the boundary, and again an estimation of the order of magnitude shows that the limit has a meaning when  $\mathbf{v}$  is a continuous differentiable

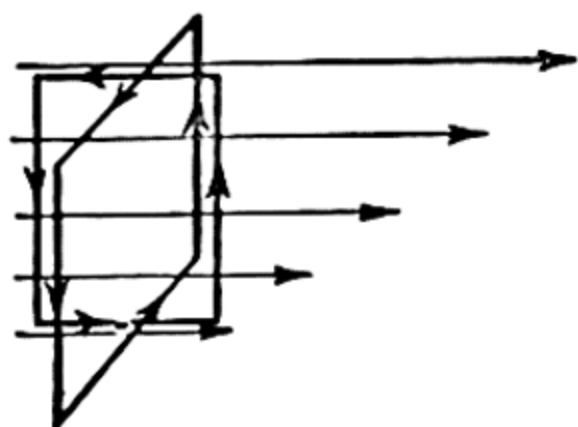


Fig. 19

point function. The value of the integral, however, depends on the orientation of the surface element, as may be shown by a simple example. Let the vector  $\mathbf{v}$  have the same direction at all points in the region, but let its magnitude, represented by the length of the arrows in fig. 19, vary in a direction at right angles to the direction of  $\mathbf{v}$ . If the surface

element is placed so that its normal is in the direction of  $\mathbf{v}$ , then  $\mathbf{v}$  is perpendicular to  $d\mathbf{s}$  along the entire circuit, and the integral vanishes. On the other hand, if the normal is perpendicular to  $\mathbf{v}$ , then the net contribution of the edges perpendicular to  $\mathbf{v}$  is zero, but that of the edges parallel to  $\mathbf{v}$  is not, since the magnitude of  $\mathbf{v}$  was taken to be different at the latter edges. The value of the limit

$$L_n = \lim \frac{1}{\Delta S} \oint \mathbf{v} d\mathbf{s} \quad . \quad . \quad (43)$$

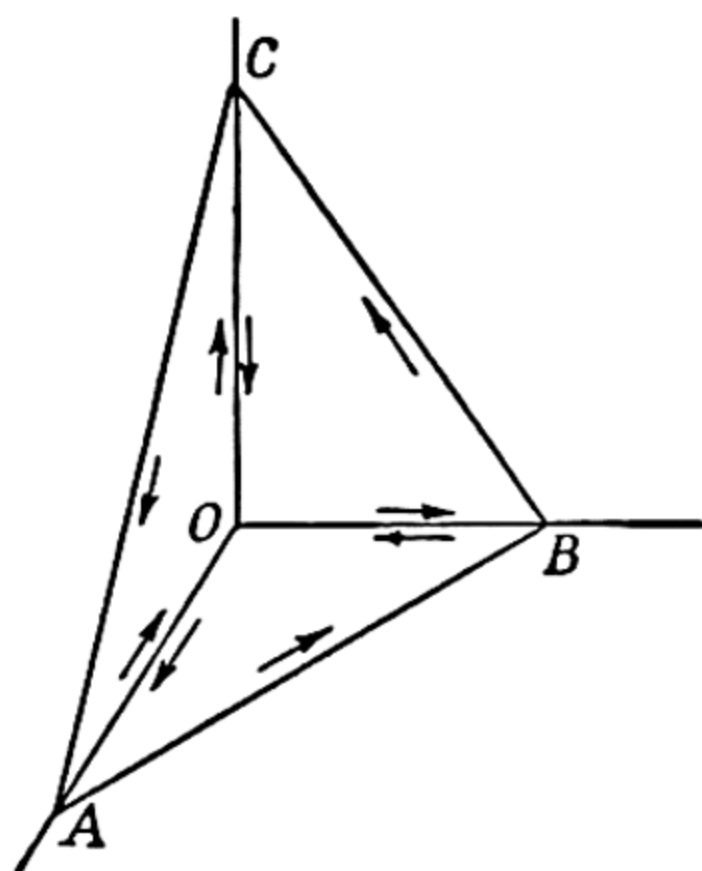


Fig. 20

is thus a function of the direction of the normal to the surface element. We now show that we can give the value of this limit for any direction, if we know the value for any three non-coplanar directions. For convenience we take these three directions to be the co-ordinate axes. As surface element we choose a small triangle  $\Delta S = ABC$ , whose vertices lie in the axes

(fig. 20). Instead of passing around the triangle  $ABC$ , we could describe, in turn, the triangles  $OBC = \Delta S_x$ ,  $OCA = \Delta S_y$  and  $OAB = \Delta S_z$ . We can calculate the limit

$$L_x = \frac{1}{\Delta S_x} \int_{OBC} \mathbf{v} d\mathbf{s}, \text{ \&c.} \quad . \quad . \quad . \quad (43')$$

for each. In adding the line integrals along the partial triangles, each axis will be covered twice, in opposite directions. There remains only the part contributed by the sides  $AB$ ,  $BC$  and  $CA$ , so that the relationship

$$L\Delta S = L_x\Delta S_x + L_y\Delta S_y + L_z\Delta S_z \quad . \quad . \quad . \quad (44)$$

is satisfied. On the other hand, from § 4 (p. 14),

$$\Delta \mathbf{S} = \mathbf{n}\Delta S = i\Delta S_x + j\Delta S_y + k\Delta S_z. \quad . \quad . \quad (45)$$

The unit normal vector to  $\Delta S$  is denoted here by  $\mathbf{n}$ , and is directed toward the outside of the tetrahedron, while  $i, j, k$  point inward. If  $\cos \alpha, \cos \beta, \cos \gamma$  are the direction cosines of the normals  $\mathbf{n}$ , this equation is, in terms of components,

$$\Delta S_x = \Delta S \cos \alpha, \quad \Delta S_y = \Delta S \cos \beta, \quad \Delta S_z = \Delta S \cos \gamma.$$

Putting these values in (44) gives

$$L_n = L_x \cos \alpha + L_y \cos \beta + L_z \cos \gamma. \quad . \quad . \quad (46)$$

The sum on the right may be interpreted as the scalar product of the unit normal vector  $\mathbf{n}$  with a vector whose components are  $L_x, L_y, L_z$ . This vector is called "the curl of  $\mathbf{v}$ " (written:  $\text{curl } \mathbf{v}$ ).\*

We have therefore

$$\text{curl } \mathbf{v} = iL_x + jL_y + kL_z. \quad . \quad . \quad . \quad (47)$$

It follows that

$$L\Delta S = \oint \mathbf{v} ds = \mathbf{n}(\text{curl } \mathbf{v})\Delta S = (\text{curl } \mathbf{v})\Delta \mathbf{S}. \quad . \quad (48)$$

Just as we obtained Gauss's Theorem, when considering the divergence, by passing from infinitesimal volume elements to finite volumes, so we obtain here a new and important relationship by considering a portion of a surface bounded by a curve. This curve need not be plane, but is to be traversed in a given sense (fig. 21). Let the surface be subdivided into infinitely small elements. The boundary of each element is to be traversed in the same sense as the outer curve. Then, for each element,



Fig. 21

$$\oint \mathbf{v} ds = (\text{curl } \mathbf{v})\Delta \mathbf{S}_i.$$

If we add these equations for the separate elements, the contributions from the inner dividing lines cancel out on the left side of the resulting equation, since each segment of these lines is traversed twice, in opposite directions. The left member is then merely the integral along the bounding curve, and we obtain the important formula called *Stokes's Theorem*:

$$\oint_C \mathbf{v} ds = \int \text{curl } \mathbf{v} d\mathbf{S}. \quad . \quad . \quad . \quad (49)$$

*The line integral of the vector  $\mathbf{v}$ , taken over a closed curve  $C$ , is equal to the surface integral of the curl of  $\mathbf{v}$ , taken over any surface having  $C$  as a boundary.*

\* The vector is sometimes called "the rotation of  $\mathbf{v}$ ", and written:  $\text{rot } \mathbf{v}$ .

There follows at once an important property of the vector,  $\text{curl } \mathbf{v}$ , its divergence is zero. If we put  $\text{curl } \mathbf{v} = \mathbf{w}$ , then Gauss's Theorem may be applied to a closed surface, and

$$\oint_S \mathbf{w} d\mathbf{S} = \int \text{div } \mathbf{w} d\tau = \int \text{div } \text{curl } \mathbf{v} d\tau.$$

Let this closed surface be divided into two parts by a closed curve (fig. 22). Then

$$\int \text{div } \text{curl } \mathbf{v} d\tau = \int_{S_1} \mathbf{w} d\mathbf{S} + \int_{S_2} \mathbf{w} d\mathbf{S} = \int_{S_1} \text{curl } \mathbf{v} d\mathbf{S} + \int_{S_2} \text{curl } \mathbf{v} d\mathbf{S}.$$

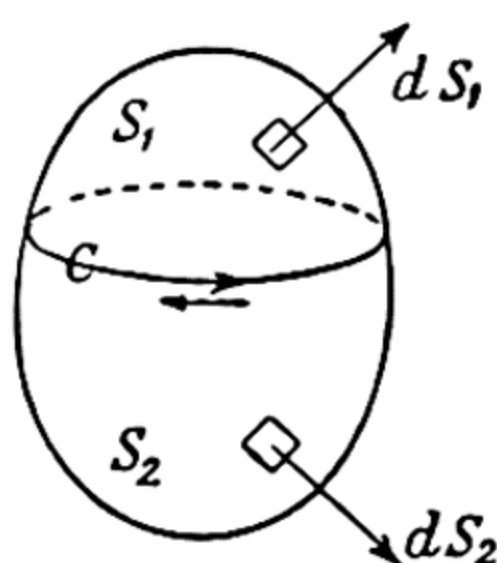


Fig. 22

If, now, we apply Stokes's Theorem to each of the two parts, we must remember that in forming the integral over the entire surface, all the normals are directed outward; but in the application of this theorem, the direction of the normals is determined by the direction in which the curve  $C$  is traversed. If we agree to take  $d\mathbf{S}$  always toward the outside, then the sign must be reversed in transforming the second integral. We thus have

$$\int \text{div } \text{curl } \mathbf{v} d\tau = \oint_C \mathbf{v} d\mathbf{s} - \oint_C \mathbf{v} d\mathbf{s} = 0.$$

Since this equation holds for any volume, it follows that

$$\text{div } \text{curl } \mathbf{v} = 0. \quad . \quad . \quad . \quad . \quad . \quad (50)$$

Further, the vanishing of the line integral of the gradient of a scalar point function, taken over a closed curve [(38') (p. 22)], may be expressed in the following way by applying Stokes's Theorem:

$$\text{curl } \text{grad } u = 0. \quad . \quad . \quad (51)$$

It remains to calculate the components of the vector  $\text{curl } \mathbf{v}$  from the components of the vector  $\mathbf{v}$ . From (47) (p. 27) and (43') (p. 26), we obtain the  $x$ -component of  $\text{curl } \mathbf{v}$  by traversing a surface element which lies

in the  $yz$ -plane. Let this element have the form of a square of side  $2h$  (cf. fig. 23), with sides parallel to the axes. Only the  $y$ -component of  $\mathbf{v}$  contributes to the integral along the horizontal sides, only the  $z$ -component contributes on the vertical sides. If we let the subscript 0 refer to the values at the centre of the square, then  $v_y$  along  $AB$  has the value

$$(v_y)_0 - \left( \frac{\partial v_y}{\partial z} \right)_0 h + \left( \frac{\partial v_y}{\partial y} \right)_0 y.$$

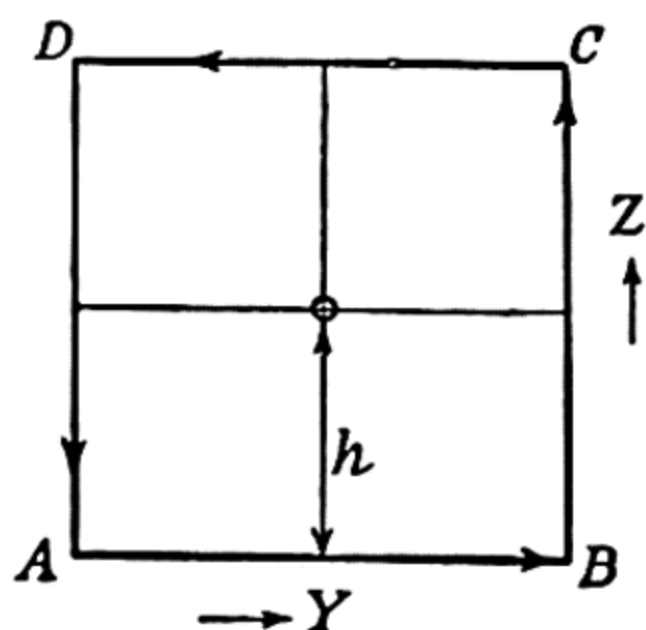


Fig. 23



Along  $CD$ :  $(v_y)_0 + \left(\frac{\partial v_y}{\partial z}\right)_0 h + \left(\frac{\partial v_y}{\partial y}\right)_0 y.$

The contribution of the side  $AB$  thus is

$$\int_A^B v ds = \int_{-h}^{+h} v_y dy = 2h(v_y)_0 - 2h^2 \left(\frac{\partial v_y}{\partial z}\right)_0.$$

In the same way, the side  $CD$  yields

$$\int_C^D v ds = - \int_{-h}^{+h} v_y dy = -2h(v_y)_0 - 2h^2 \left(\frac{\partial v_y}{\partial z}\right)_0.$$

Together, these two sides give

$$\int_A^B v ds + \int_C^D v ds = -4h^2 \left(\frac{\partial v_y}{\partial z}\right)_0.$$

Similarly, the two vertical sides give

$$\int_B^O v ds + \int_D^A v ds = 4h^2 \left(\frac{\partial v_z}{\partial y}\right)_0.$$

Since the area of the square is  $4h^2$ , division by this quantity yields \*

$$(\text{curl } \mathbf{v})_x = \frac{\partial v_z}{\partial y} - \frac{\partial v_y}{\partial z},$$

and, in like manner,

$$(\text{curl } \mathbf{v})_y = \frac{\partial v_x}{\partial z} - \frac{\partial v_z}{\partial x},$$

$$(\text{curl } \mathbf{v})_z = \frac{\partial v_y}{\partial x} - \frac{\partial v_x}{\partial y}.$$

Hence

$$\text{curl } \mathbf{v} = \left(\frac{\partial v_z}{\partial y} - \frac{\partial v_y}{\partial z}\right) i + \left(\frac{\partial v_x}{\partial z} - \frac{\partial v_z}{\partial x}\right) j + \left(\frac{\partial v_y}{\partial x} - \frac{\partial v_x}{\partial y}\right) k. \quad (47')$$

On the understanding that the "product" of the differentiating operator  $\partial/\partial x$  with a quantity  $u$  means  $\partial u/\partial x$ , then the last equation may be written as a determinant:

$$\text{curl } \mathbf{v} = \begin{vmatrix} i & j & k \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ v_x & v_y & v_z \end{vmatrix}. \quad \dots \dots (52)$$

**Ex. 16.** Calculate  $\text{curl } \mathbf{r}$ .

\* The subscripts are omitted; the derivatives refer, as before, to the centre of the element.

10. The Operator  $\nabla$ .

If we compare the three formulæ

$$(36) \text{ (p. 21): } \text{grad } u = i \frac{\partial u}{\partial x} + j \frac{\partial u}{\partial y} + k \frac{\partial u}{\partial z},$$

$$(42) \text{ (p. 25): } \text{div } \mathbf{v} = \frac{\partial v_x}{\partial x} + \frac{\partial v_y}{\partial y} + \frac{\partial v_z}{\partial z},$$

$$(47') \text{ (p. 29): } \text{curl } \mathbf{v} = i \left( \frac{\partial v_z}{\partial y} - \frac{\partial v_y}{\partial z} \right) + j \left( \frac{\partial v_x}{\partial z} - \frac{\partial v_z}{\partial x} \right) + k \left( \frac{\partial v_y}{\partial x} - \frac{\partial v_x}{\partial y} \right),$$

we recognize that we can express all of them symbolically in terms of a vector whose components are  $\partial/\partial x$ ,  $\partial/\partial y$ ,  $\partial/\partial z$ , on the principle just explained (cf. the last sentence of the preceding section). This symbolical vector is denoted by  $\nabla$ . It was introduced by Sir William Hamilton, and is called "nabla", after an ancient Assyrian harp whose form  $\nabla$  resembles: \*

$$\nabla = i \frac{\partial}{\partial x} + j \frac{\partial}{\partial y} + k \frac{\partial}{\partial z}. \quad . \quad . \quad . \quad . \quad . \quad (53)$$

If we multiply a scalar quantity  $u$  with this vector operator, we obtain

$$\nabla u = i \frac{\partial u}{\partial x} + j \frac{\partial u}{\partial y} + k \frac{\partial u}{\partial z} = \text{grad } u. \quad . \quad . \quad (54)$$

If we form the scalar product of the nabla operator with a vector  $\mathbf{v}$ , we obtain, according to the definition of scalar product, the sum of the products of corresponding components:

$$\nabla \cdot \mathbf{v} = \frac{\partial v_x}{\partial x} + \frac{\partial v_y}{\partial y} + \frac{\partial v_z}{\partial z} = \text{div } \mathbf{v}. \quad . \quad . \quad . \quad (55)$$

Finally, equation (52) (p. 29) represents the vector product of the nabla operator with  $\mathbf{v}$ :

$$[\nabla \mathbf{v}] = \text{curl } \mathbf{v}. \quad . \quad . \quad . \quad . \quad . \quad (56)$$

From the formal correspondence of the operations grad, div, and curl with multiplication, it follows at once that these operations are distributive with respect to addition. That is,

$$\nabla(\mathbf{v}_1 + \mathbf{v}_2) = \nabla \mathbf{v}_1 + \nabla \mathbf{v}_2, \text{ \&c.} \quad . \quad . \quad . \quad (57)$$

By starting with the definition of divergence, and generalizing, we obtain a definition of the nabla operator which is independent of the co-ordinate system:

\* The name "nabla" was used by *Heaviside*. This operator is also frequently called "del", and sometimes "atled" (the word "delta" reversed).



that is,

$$\left. \begin{aligned} \operatorname{div} \boldsymbol{v} &= \nabla \cdot \boldsymbol{v} = \lim \frac{1}{\Delta \tau} \oint d\boldsymbol{S} \cdot \boldsymbol{v}, \\ \nabla u &= \lim \frac{1}{\Delta \tau} \oint d\boldsymbol{S} u, \\ \operatorname{grad} u &= \nabla u = \lim \frac{1}{\Delta \tau} \oint d\boldsymbol{S} u, \\ \operatorname{curl} \boldsymbol{v} &= [\nabla \times \boldsymbol{v}] = \lim \frac{1}{\Delta \tau} \oint [d\boldsymbol{S} \times \boldsymbol{v}]. \end{aligned} \right\} \dots \dots (58)$$

We shall illustrate the method of proving these formulæ by taking the case of the gradient. Since the form of the volume element is entirely at our disposal, we take a small cylinder, the generating lines of which are in the direction of  $\operatorname{grad} u$  and are therefore normal to the level surface (fig. 24). Apart from terms of higher order, the value of  $u$  over the base of the cylinder is  $u_0 - h |\operatorname{grad} u|$ ; at the top of the cylinder, the value is  $u_0 + h |\operatorname{grad} u|$ . It is to be remembered that, in forming the surface integral, the elements  $d\boldsymbol{S}$  must always point outward from the elementary volume. The lateral surface of the cylinder does not contribute to the integral, since  $u$  has a constant value in each transverse section; and to each surface element there corresponds another which is equal and opposite. Also, we obtain the contributions of the ends of the cylinder by multiplying the area  $\Delta S$  by the difference in the values of  $u$ , and by expressing the direction of  $d\boldsymbol{S}$  by means of  $\operatorname{grad} u$ :

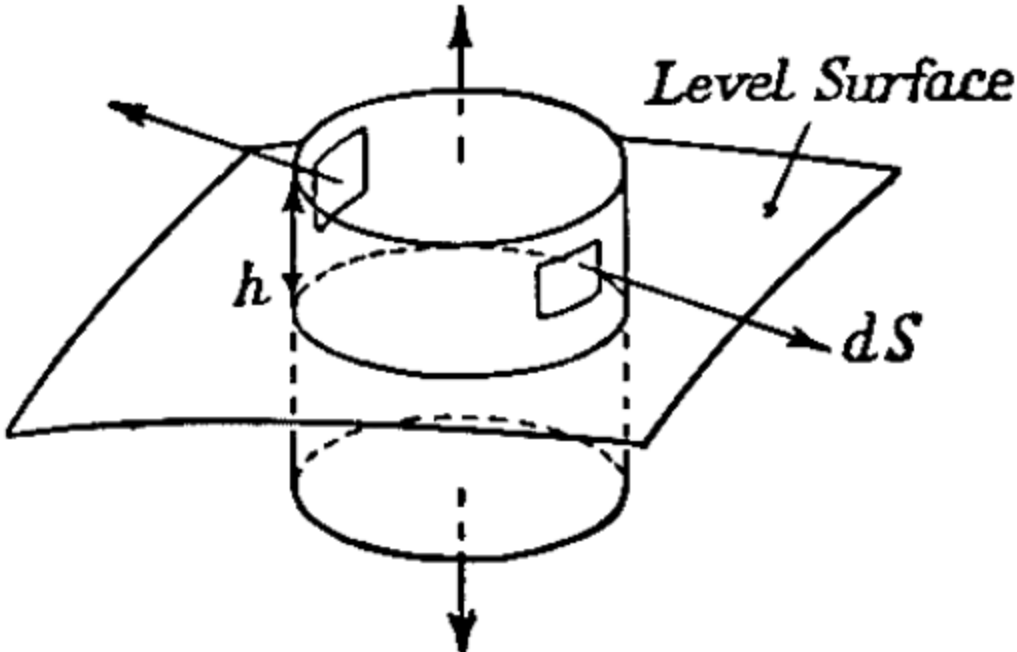


Fig. 24

$$\oint d\boldsymbol{S} u = (\Delta S) (2h \operatorname{grad} u) = \Delta \tau \operatorname{grad} u.$$

Dividing by the volume, and passing to the limit, we obtain the third equation of (58).

These formulæ yield important generalizations of Gauss's Theorem, the proof of which is analogous to that given above:

$$\oint d\boldsymbol{S} u = \int \operatorname{grad} u d\tau. \dots \dots (59)$$

$$\oint [d\boldsymbol{S} \times \boldsymbol{v}] = \int \operatorname{curl} \boldsymbol{v} d\tau. \dots \dots (60)$$

Theorems (59) and (60) are also easily proved by writing them out in their scalar forms (which are often useful) and using (41). For (41) can be written

$$\int (lA + mB + nC) dS = \int \left( \frac{\partial A}{\partial x} + \frac{\partial B}{\partial y} + \frac{\partial C}{\partial z} \right) d\tau,$$

where  $A, B, C$  are scalar functions, and  $l, m, n$  are the direction cosines of the outward normal to  $S$ . Since  $A, B, C$  are independent, it follows that

$$\int lA dS = \int \frac{\partial A}{\partial x} d\tau, \quad \int mB dS = \int \frac{\partial B}{\partial y} d\tau, \quad \int nC dS = \int \frac{\partial C}{\partial z} d\tau.$$

These formulæ are closely connected with Green's Theorem (p. 270). They are equivalent to (59), and also lead at once to (60), which involves three equations of the type

$$\int (mv_z - nv_y) dS = \int \left( \frac{\partial v_z}{\partial y} - \frac{\partial v_y}{\partial z} \right) d\tau.$$

### 11. Calculation of the Gradient in a Vector Field; Fundamental Principles of Tensor Analysis.

In § 7 (p. 20) the question "How does a scalar point function change with a displacement  $dS$ ?" was treated. This question may be extended to vector fields, and we may inquire "How does the vector  $\mathbf{v}$  change with a displacement  $dS$ ?" If we start with the components of the vector, this question may be referred back to the case of a scalar point function. If a vector field be given, then at each point in space we have a field of three scalars, viz. the three components. But, according to § 7 (p. 21), the change in the components may be written down at once:

$$\left. \begin{aligned} dv_x &= dS \operatorname{grad} v_x = \frac{\partial v_x}{\partial x} dx + \frac{\partial v_x}{\partial y} dy + \frac{\partial v_x}{\partial z} dz, \\ dv_y &= dS \operatorname{grad} v_y = \frac{\partial v_y}{\partial x} dx + \frac{\partial v_y}{\partial y} dy + \frac{\partial v_y}{\partial z} dz, \\ dv_z &= dS \operatorname{grad} v_z = \frac{\partial v_z}{\partial x} dx + \frac{\partial v_z}{\partial y} dy + \frac{\partial v_z}{\partial z} dz. \end{aligned} \right\} \quad (61)$$

and so

$$\begin{aligned} d\mathbf{v} &= i dv_x + j dv_y + k dv_z \\ &= dS \operatorname{grad} v_x \cdot i + dS \operatorname{grad} v_y \cdot j + dS \operatorname{grad} v_z \cdot k. \end{aligned} \quad (61')$$

Merely as an abbreviation we write

$$d\mathbf{v} = dS \operatorname{grad} \cdot \mathbf{v}, \quad \dots \dots \dots (62)$$

which is read: " $dS$ -vector gradient of  $\mathbf{v}$ ."

In order to be able to calculate  $d\mathbf{v}$ , three vectors (or nine scalars) must be known. We shall see that such forms, consisting of nine numbers, have also a physical meaning. From (61) we see that the components of the vector  $d\mathbf{v}$  are linear functions of the components of the vector  $dS$ , the coefficients of these functions being the partial derivatives  $\partial v_x / \partial x$ ,  $\partial v_x / \partial y$ , &c.

Let us now investigate such linear vector functions in detail. As

in (61), let the components of a vector  $w$  be linear functions of the components of a vector  $v$ :

$$\left. \begin{aligned} w_x &= a_{11} v_x + a_{12} v_y + a_{13} v_z, \\ w_y &= a_{21} v_x + a_{22} v_y + a_{23} v_z, \\ w_z &= a_{31} v_x + a_{32} v_y + a_{33} v_z. \end{aligned} \right\} \quad \cdot \quad \cdot \quad \cdot \quad (63)$$

In this manner the vector  $w$  is placed in one-to-one correspondence with the vector  $v$ . The scheme of coefficients  $a_{ik}$  has thus an independent meaning if this correspondence is such that the passage from  $v$  to  $w$  is independent of the particular co-ordinate system in which the vectors are resolved into components. In this case, the coefficients  $a_{ik}$  are said to be the components of a Tensor.\*

A vector is a physically determined directed quantity in space, but the values of the components depend on the choice of axes. In the  $i, j, k$  system,  $v_x = v i, \quad v_y = v j, \quad v_z = v k$ .

For a system  $i', j', k'$ , turned with respect to this system,

$$\left. \begin{aligned} i' &= \beta_{11} i + \beta_{12} j + \beta_{13} k, \\ j' &= \beta_{21} i + \beta_{22} j + \beta_{23} k, \\ k' &= \beta_{31} i + \beta_{32} j + \beta_{33} k, \end{aligned} \right\} \quad \cdot \quad \cdot \quad \cdot \quad (64)$$

where, since  $i', j', k'$  again form a rectangular system, the  $\beta_{ik}$  represent the direction cosines of the new axes referred to the old; e.g.  $\beta_{ik}$  are the direction cosines of the new axes with respect to the old. Then, as may be verified by forming the scalar products  $i' j' = 0, i' i' = 1$ , &c.,

$$\begin{aligned} \beta_{11}^2 + \beta_{12}^2 + \beta_{13}^2 &= 1, \\ \beta_{11}\beta_{21} + \beta_{12}\beta_{22} + \beta_{13}\beta_{23} &= 0; \end{aligned}$$

in general,

$$\left. \begin{aligned} \Sigma \beta_{ik} \beta_{jk} &= 1 \text{ for } i = j \\ &= 0 \text{ for } i \neq j. \end{aligned} \right\} \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad (65)$$

Thus the old axes are expressed in terms of the new:

$$\left. \begin{aligned} i &= \beta_{11} i' + \beta_{21} j' + \beta_{31} k', \\ j &= \beta_{12} i' + \beta_{22} j' + \beta_{32} k', \\ k &= \beta_{13} i' + \beta_{23} j' + \beta_{33} k'. \end{aligned} \right\} \quad \cdot \quad \cdot \quad \cdot \quad (66)$$

The components of  $v$  become, in the new co-ordinate system,

$$\left. \begin{aligned} v_x &= v i' = \beta_{11} v i + \beta_{12} v j + \beta_{13} v k = \beta_{11} v_x + \beta_{12} v_y + \beta_{13} v_z, \\ v_y &= v j' = \beta_{21} v i + \beta_{22} v j + \beta_{23} v k = \beta_{21} v_x + \beta_{22} v_y + \beta_{23} v_z, \\ v_z &= v k' = \beta_{31} v i + \beta_{32} v j + \beta_{33} v k = \beta_{31} v_x + \beta_{32} v_y + \beta_{33} v_z, \end{aligned} \right\} \quad (67)$$

i.e. the vector components transform in the same way as the axes. We thus have also

$$\left. \begin{aligned} v_x &= \beta_{11} v_{x'} + \beta_{21} v_{y'} + \beta_{31} v_{z'}, \\ v_y &= \beta_{12} v_{x'} + \beta_{22} v_{y'} + \beta_{32} v_{z'}, \\ v_z &= \beta_{13} v_{x'} + \beta_{23} v_{y'} + \beta_{33} v_{z'}. \end{aligned} \right\} \quad \cdot \quad \cdot \quad \cdot \quad (68)$$

\* Also called in German works an *Affinor*, the term originating in the fact that for  $v = r$  the  $a_{ik}$  give an affine transformation of space (or homogeneous strain).



How must the tensor components  $a_{ik}$  transform, if a change of co-ordinate system is not to alter the relationship of  $\mathbf{w}$  to  $\mathbf{v}$ ? The following equations must hold:

$$\left. \begin{aligned} w_x &= a_{11}v_x + a_{12}v_y + a_{13}v_z, \\ w_y &= a_{21}v_x + a_{22}v_y + a_{23}v_z, \\ w_z &= a_{31}v_x + a_{32}v_y + a_{33}v_z, \end{aligned} \right\} \quad . \quad . \quad . \quad (69)$$

and

$$\left. \begin{aligned} w_{x'} &= a_{11}'v_{x'} + a_{12}'v_{y'} + a_{13}'v_{z'}, \\ w_{y'} &= a_{21}'v_{x'} + a_{22}'v_{y'} + a_{23}'v_{z'}, \\ w_{z'} &= a_{31}'v_{x'} + a_{32}'v_{y'} + a_{33}'v_{z'}. \end{aligned} \right\} \quad . \quad . \quad . \quad (70)$$

In addition,  $w_x, w_y, w_z$  and  $w_{x'}, w_{y'}, w_{z'}$  must be so related that the same vector  $\mathbf{w}$  is determined by both triads. If we multiply the first of the unaccented equations by  $\beta_{11}$ , the second by  $\beta_{12}$ , the third by  $\beta_{13}$ , and add, we obtain, by (67),

$$w_{x'} = (\beta_{11}a_{11} + \beta_{12}a_{21} + \beta_{13}a_{31})v_x + (\beta_{11}a_{12} + \beta_{12}a_{22} + \beta_{13}a_{32})v_y + (\beta_{11}a_{13} + \beta_{12}a_{23} + \beta_{13}a_{33})v_z. \quad . \quad . \quad . \quad (71)$$

If we express the  $v_x, v_y, v_z$  here in terms of the new components, according to (68), we obtain  $w_{x'}$  as a function of  $v_{x'}, v_{y'}, v_{z'}$ . This value of  $w_{x'}$  must be identical with that defined by (70). Comparing the coefficients of  $v_{x'}, v_{y'}, v_{z'}$ , we have

$$\begin{aligned} a_{11}' &= \beta_{11}^2 a_{11} + \beta_{11}\beta_{12}a_{12} + \beta_{11}\beta_{13}a_{13} + \beta_{12}\beta_{11}a_{21} + \beta_{12}^2 a_{22} \\ &\quad + \beta_{12}\beta_{13}a_{23} + \beta_{13}\beta_{11}a_{31} + \beta_{13}\beta_{12}a_{32} + \beta_{13}^2 a_{33}. \end{aligned} \quad . \quad (72)$$

Thus the tensor components transform, under rotation of axes, in such manner that the new components become linear functions of all the old ones, the coefficients being the squares and products of the  $\beta_{ik}$ .

To find the coefficients for the tensor components  $a_{ik}'$  set  $v_{x'} = v_1'$ ,  $v_x = v_1$ , &c., and use (67) to form the product  $v_i'v_k'$  of the vector components, remembering that in expanding the brackets the order of the factors  $v_i$  and  $v_k$  must not be changed. Replacing  $v_i'v_k'$  by  $a_{ik}'$ , the coefficients of  $v_i v_k$  yield those of  $a_{ik}$ .

It is possible to represent a tensor by means of vectors. For the three components of  $\mathbf{w}$  in equation (63) (p. 33) may be written as the scalar products of the vector  $\mathbf{v}$  with the three vectors  $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$ . This amounts to regarding the three coefficients of each row as vector components:

$$\left. \begin{aligned} \mathbf{a}_i &= a_{i1}\mathbf{i} + a_{i2}\mathbf{j} + a_{i3}\mathbf{k}, \\ \mathbf{w} &= \mathbf{a}_1\mathbf{v} \cdot \mathbf{i} + \mathbf{a}_2\mathbf{v} \cdot \mathbf{j} + \mathbf{a}_3\mathbf{v} \cdot \mathbf{k}. \end{aligned} \right\} \quad . \quad . \quad . \quad (73)$$

But, as may be seen in the example of forming a gradient, the three vectors depend upon the choice of co-ordinate system, for the three vectors  $\text{grad } v_x, \text{grad } v_y$  and  $\text{grad } v_z$  are dependent upon the co-ordinate system.

Since the components of  $\mathbf{w}$  are linear homogeneous functions of the components of  $\mathbf{v}$ , according to (63) (p. 33), this functional correspondence  $\Phi$  satisfies the relationship

$$\Phi(\mathbf{v}_1 + \mathbf{v}_2) = \Phi(\mathbf{v}_1) + \Phi(\mathbf{v}_2), \quad . \quad . \quad . \quad (74)$$



i.e.  $\Phi$  is distributive with respect to addition. But this is the essential property of multiplication, since the other properties of products have already been given up in the case of the vector product. We may thus consider the vector  $\omega$  as resulting from the multiplication of the vector  $v$  with the tensor  $\Phi$ . This tensor may be represented, according to (73), by three *pairs* of vectors (*dyads*):

$$a_1 \cdot i + a_2 \cdot j + a_3 \cdot k. \quad . \quad . \quad . \quad . \quad . \quad (75)$$

The first factors in each term are called the *antecedents*, the second factors, the *consequents*; the dot does not indicate multiplication, but merely separates the factors. We understand  $\Phi v$  to mean a vector which results from scalar multiplication of the consequents with  $v$ , and  $v \Phi$  means a vector obtained from  $v$  by scalar multiplication of the antecedents with  $v$ . Thus

$$\Phi v = a_1 \cdot i v + a_2 \cdot j v + a_3 \cdot k v, \quad . \quad . \quad (76)$$

$$v \Phi = a_1 v \cdot i + a_2 v \cdot j + a_3 v \cdot k. \quad . \quad . \quad (77)$$

The two vectors are, in general, different. The tensor resulting from interchanging the antecedents and consequents is called the *conjugate tensor*  $\Phi_o$ .

$$\left. \begin{aligned} \Phi_o &= i \cdot a_1 + j \cdot a_2 + k \cdot a_3 \\ \Phi_o v &= i \cdot a_1 v + j \cdot a_2 v + k \cdot a_3 v. \end{aligned} \right\} \quad . \quad . \quad (78)$$

Since it is immaterial whether the scalar  $a_1 v$  is written before or after the vector  $i$ , comparison with (77) shows that

$$v \Phi = \Phi_o v. \quad . \quad . \quad . \quad . \quad . \quad . \quad (79)$$

In the component representation of a tensor, changing to the conjugate tensor means the interchange of rows and columns in the array of coefficients.

The dyad, indicated by writing two vectors side by side with a dot separating them,\* also has the leading property of multiplication, viz. distributivity with respect to addition:

$$(a_1 + a_2) \cdot i = a_1 \cdot i + a_2 \cdot i. \quad . \quad . \quad . \quad . \quad (80)$$

To show this it is sufficient to form the scalar product of the antecedents with a vector  $v$ , in which case

$$v(a_1 + a_2) \cdot i = v a_1 \cdot i + v a_2 \cdot i,$$

since the scalar products are distributive; the result therefore follows. We may therefore look upon the union of two vectors to form a dyad

\* Parentheses may, in some cases, be necessary for clarity, e.g. if the factors are multinomials.

as a new kind of non-commutative product, called the direct or *dyadic product*. This leads to a new interpretation of the expression  $d\mathbf{s} \text{ grad} \cdot \mathbf{v}$  with which we started. If we form the dyadic product of the operator nabla with  $\mathbf{v}$ , there results

$$\begin{aligned} \nabla \cdot \mathbf{v} &= \left( i \frac{\partial}{\partial x} + j \frac{\partial}{\partial y} + k \frac{\partial}{\partial z} \right) \cdot (v_x i + v_y j + v_z k) \\ &= \left( i \frac{\partial v_x}{\partial x} + j \frac{\partial v_x}{\partial y} + k \frac{\partial v_x}{\partial z} \right) \cdot i + \left( i \frac{\partial v_y}{\partial x} + j \frac{\partial v_y}{\partial y} + k \frac{\partial v_y}{\partial z} \right) \cdot j + \dots \\ &= \text{grad } v_x \cdot i + \text{grad } v_y \cdot j + \text{grad } v_z \cdot k. \end{aligned} \quad (81)$$

If this tensor is now multiplied by the displacement  $d\mathbf{s}$ , we obtain exactly (62), viz.

$$d\mathbf{s} \nabla \cdot \mathbf{v} = d\mathbf{s} \text{ grad} \cdot \mathbf{v}. \quad (82)$$

By dyadic multiplication of the nabla operator with the field vector  $\mathbf{v}$  we obtain a tensor called the vector gradient of  $\mathbf{v}$ . Multiplication of the displacement  $d\mathbf{s}$  with this tensor as post-factor gives the change of the vector  $\mathbf{v}$  due to the displacement  $d\mathbf{s}$ .

Let us again consider tensors in general. A tensor is said to be symmetric if

$$\Phi_o = \Phi, \text{ or, in terms of components, if } a_{ik} = a_{ki}, \quad (83)$$

and is antisymmetric (skew-symmetric) if

$$\Phi_o = -\Phi, \text{ or, in terms of components, } \left. \begin{aligned} a_{ii} &= 0, \\ a_{ik} &= -a_{ki} \end{aligned} \right\}. \quad (83')$$

By (83), the symmetric tensor has but 6 independent components, and, according to (83'), the antisymmetric tensor has only three components, i.e. as many as a vector. As a matter of fact, in three dimensions, the multiplication of a vector with an antisymmetric tensor always may be replaced by forming the vector product with a vector. In physics, symmetric tensors are specially important. They may be represented by a surface of the second degree. The vectors  $\mathbf{v}$  are drawn as radius vectors  $\mathbf{r}$  from a fixed origin. By means of the tensor  $\Theta$ , a vector  $\mathbf{w} = \mathbf{r}'$  is allocated to each vector  $\mathbf{r}$ . Let the scalar product  $\mathbf{r}\mathbf{r}'$  be designated by  $p$ . Then

$$\mathbf{r}\mathbf{r}' = p = \mathbf{r}(\Theta \mathbf{r}) = a_{11}x^2 + a_{12}xy + a_{13}xz + a_{21}yx + a_{22}y^2 + \dots$$

Because of the proviso that  $\Theta$  is to be a symmetric tensor, we have

$$p = \mathbf{r}(\Theta \mathbf{r}) = a_{11}x^2 + a_{22}y^2 + a_{33}z^2 + 2a_{12}xy + 2a_{23}yz + 2a_{31}zx, \quad (84)$$

from which there follows

$$\left. \begin{aligned} r_x' &= a_{11}x + a_{12}y + a_{13}z = \frac{1}{2} \frac{\partial p}{\partial x}, \\ r' &= \frac{1}{2} \text{grad } p. \end{aligned} \right\} \quad . . . \quad (85)$$

This gives a graphical construction for  $r'$ . By (85),  $r'$  has the direction of the normal to the surface  $p = \text{const.}$ , at the end point  $P$  of  $r$ . If we draw the surface  $p = 1$ , then, according to (84), the *magnitude* of  $r'$  is the reciprocal of the projection of  $r$  on the direction of the normal at  $P$  (cf. fig. 25).

The surface  $p = 1$  is called the *tensor ellipsoid*, although without further limitation on the values of the coefficients, it might be any second degree surface. The representation by means of a second degree surface is seen to be possible, since both a surface of this kind and a *symmetric* tensor have six independent coefficients.

Further, for every surface of the second degree there is one specially appropriate co-ordinate system, viz. that of the principal axes. In this system of axes the surface has the equation

$$p = a_I x'^2 + a_{II} y'^2 + a_{III} z'^2 \quad . . . \quad (84')$$

and the tensor, referred to these new  $i', j', k'$  axes, becomes simply

$$\Theta = a_I i' \cdot i' + a_{II} j' \cdot j' + a_{III} k' \cdot k'. \quad . . \quad (75')$$

The components of  $w$  are thus given by

$$w_x' = a_I v_x', \quad w_y' = a_{II} v_y'; \quad w_z' = a_{III} v_z'. \quad . . \quad (86)$$

*Ex. 17.* Write the *Frenet* formulæ [equns. (32), p. 19, and (35), p. 20], using the vector gradient.

*Ex. 18.* Determine the directions of the principal axes of a tensor ellipsoid by using the relation that  $r' = \lambda r$  in these directions.

## 12. Calculation of more complicated Vector Differential Expressions with the help of the Nabla Operator.

The operations grad, div, curl and grad may be performed readily, in the case of products of scalar or vector point functions, with the help of the  $\nabla$ -operator. It is only necessary to remember that  $\nabla$  is a sign of differentiation as well as a vector. Everything standing to the right of such a sign is subject to the process of differentiation. In ordinary differentiation, the quantities to be held constant may always be placed to the left of the differentiating symbol. But since the

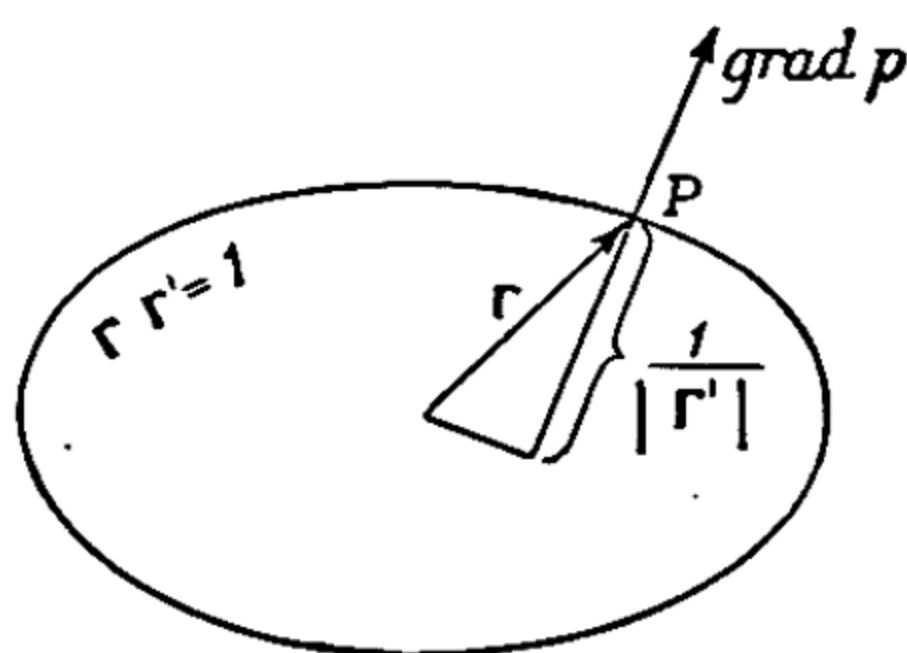


Fig. 25



vector product, and especially the dyadic product, of two vectors are not commutative—though of course in a vector product we can always change the order by changing the sign—we shall indicate by means of the subscript  $c$  the constancy of a factor on the right of the sign of differentiation. This will permit us to leave constants on the right side of the sign  $\nabla$ . By following a limiting process entirely analogous to that for the ordinary differentiation of products, it may be shown that, in general, a space derivative (grad, div, curl, **grad**) of a product is equal to the sum of the corresponding expressions for products in which only one factor is variable. Thus we obtain the following practical rule for the space differentiation of products: *Using the  $\nabla$ -symbol, write the differential expression as the sum of derivatives in which only one factor is subject to differentiation. Transform these by the rules of vector algebra in such a way that all quantities not to be differentiated are on the left of the sign of differentiation, and interpret the resulting expressions in terms of the special symbols, grad, div, &c.*

This method is easily understood in the case of the more important and frequently occurring formulæ. These expressions will be computed below.

(a) *Products of scalar point functions*

$$\begin{aligned} 1. \quad \text{grad } uvw &= \nabla uv_c w_c + \nabla u_c v w_c + \nabla u_c v_c w \\ &= vw \text{ grad } u + wu \text{ grad } v + uv \text{ grad } w. \quad . \quad . \quad (87) \end{aligned}$$

(b) *Product of a scalar point function with a vectorial point function*

$$2. \quad \text{div } uv = \nabla uv_c + \nabla u_c v = v \text{ grad } u + u \text{ div } v. \quad . \quad (88)$$

$$3. \quad \text{curl } uv = [\nabla uv_c] + [\nabla u_c v] = [\text{grad } u, v] + u \text{ curl } v. \quad (89)$$

$$4. \quad \text{grad} \cdot uv = \nabla \cdot uv_c + \nabla \cdot u_c v = \text{grad } u \cdot v + u \text{ grad} \cdot v. \quad (90)$$

(c) *Products of vectorial point functions*

$$\text{grad } v w = \nabla \cdot v w_c + \nabla \cdot v_c w.$$

$$\begin{aligned} \text{Now} \quad [v \text{ curl } w] &= [v [\nabla w]] = \nabla \cdot v_c w - v \nabla \cdot w \\ &= \text{grad } v_c w - v \text{ grad} \cdot w; \end{aligned}$$

thus

$$\text{grad } v w = [v \text{ curl } w] + v \text{ grad} \cdot w + [w \text{ curl } v] + w \text{ grad} \cdot v. \quad (91)$$

$$\begin{aligned} \text{div}[v w] &= \nabla[v w_c] + \nabla[v_c w] = w[\nabla v] - v[\nabla w] \\ &= w \text{ curl } v - v \text{ curl } w. \quad . \quad . \quad . \quad . \quad . \quad (92) \end{aligned}$$



$$\begin{aligned}\text{curl}[\boldsymbol{v}\boldsymbol{w}] &= [\nabla[\boldsymbol{v}\boldsymbol{w}_z]] + [\nabla[\boldsymbol{v}_z\boldsymbol{w}]] \\ &= \boldsymbol{w} \text{grad} \cdot \boldsymbol{v} - \boldsymbol{w} \text{div} \boldsymbol{v} - \boldsymbol{v} \text{grad} \cdot \boldsymbol{w} + \boldsymbol{v} \text{div} \boldsymbol{w}. \quad (93)\end{aligned}$$

(d) *Second derivatives*

$$\begin{aligned}\text{div grad } u &= \nabla \nabla u = \frac{\partial}{\partial x} \frac{\partial u}{\partial x} + \frac{\partial}{\partial y} \frac{\partial u}{\partial y} + \frac{\partial}{\partial z} \frac{\partial u}{\partial z} \\ &= \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} = \Delta u. \quad . \quad . \quad . \quad (94)\end{aligned}$$

The scalar product of the  $\nabla$ -operator with itself is therefore equal to the Laplacian operator  $\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$ , denoted by  $\Delta$ . Note, however that the rule for products stated in § 12, p. 38, does not apply to  $\Delta$ , as the following example shows:

$$\begin{aligned}\text{div grad } uv &= \text{div}(u \text{grad } v + v \text{grad } u) = u \Delta v + v \Delta u \\ &\quad + 2 \text{grad } u \text{grad } v.\end{aligned}$$

Again, we have  $\text{curl grad } u = [\nabla \nabla u] = 0. \quad . \quad . \quad . \quad (49)$

This result, already obtained in § 9 (p. 28), follows purely formally here from the rule of vector algebra which states that a product of the form  $[\boldsymbol{A}\boldsymbol{A}\boldsymbol{a}]$  vanishes. Similar remarks hold for the relationship

$$\text{div curl } \boldsymbol{v} = \nabla[\nabla \boldsymbol{v}] = 0, \quad . \quad . \quad . \quad (48)$$

also obtained in § 9 (p. 28).

On the other hand, there is no simpler vector expression for  $\text{grad div } \boldsymbol{v}$ . In rectangular co-ordinates, this expression signifies the vector

$$\begin{aligned}i \frac{\partial}{\partial x} \text{div } \boldsymbol{v} + j \frac{\partial}{\partial y} \text{div } \boldsymbol{v} + k \frac{\partial}{\partial z} \text{div } \boldsymbol{v} &= i \left( \frac{\partial^2 v_x}{\partial x^2} + \frac{\partial^2 v_y}{\partial x \partial y} + \frac{\partial^2 v_z}{\partial x \partial z} \right) \\ &\quad + j \left( \frac{\partial^2 v_x}{\partial x \partial y} + \frac{\partial^2 v_y}{\partial y^2} + \frac{\partial^2 v_z}{\partial y \partial z} \right) + k \left( \frac{\partial^2 v_x}{\partial x \partial z} + \frac{\partial^2 v_y}{\partial y \partial z} + \frac{\partial^2 v_z}{\partial z^2} \right). \quad (95)\end{aligned}$$

With the help of the expression for  $\text{grad div } \boldsymbol{v}$ , the expression  $\text{curl curl } \boldsymbol{v}$  may be transformed in a manner which finds important application in electrodynamics:

$$\text{curl curl } \boldsymbol{v} = [\nabla[\nabla \boldsymbol{v}]] = \nabla \cdot \nabla \boldsymbol{v} - \nabla \nabla \cdot \boldsymbol{v} = \text{grad div } \boldsymbol{v} - \nabla \nabla \cdot \boldsymbol{v}. \quad (96)$$

Here we have the scalar product of  $\nabla$  with itself standing before a vector. If this vector is written in terms of its components, the complete expression  $\text{curl curl } \boldsymbol{v}$  may be calculated with little difficulty;

the computation shows that the result is  $\text{grad div } \mathbf{v}$  minus a vector with components as follows:

$$\begin{aligned} \nabla \nabla \cdot \mathbf{v} &= i \left( \frac{\partial^2 v_x}{\partial x^2} + \frac{\partial^2 v_x}{\partial y^2} + \frac{\partial^2 v_x}{\partial z^2} \right) + j \left( \frac{\partial^2 v_y}{\partial x^2} + \frac{\partial^2 v_y}{\partial y^2} + \frac{\partial^2 v_y}{\partial z^2} \right) \\ &+ k \left( \frac{\partial^2 v_z}{\partial x^2} + \frac{\partial^2 v_z}{\partial y^2} + \frac{\partial^2 v_z}{\partial z^2} \right) = i \Delta v_x + j \Delta v_y + k \Delta v_z = \Delta \mathbf{v}. \quad (97) \end{aligned}$$

Thus the rule  $\nabla \nabla = \Delta$ , which held for simple point functions, is true here also. Hence we have the important relation

$$\text{curl curl } \mathbf{v} = \text{grad div } \mathbf{v} - \Delta \mathbf{v}. \quad . . . \quad (98)$$

*Ex. 19.* Calculate the following surface integrals for constant vector  $\mathbf{i}$ :

$$(a) \oint (\mathbf{i} \cdot \mathbf{r}) \cdot d\mathbf{S}; \quad (b) \oint \mathbf{r} \cdot \mathbf{i} d\mathbf{S}.$$

*Ex. 20.* Calculate  $\Delta \frac{1}{r}$ .

*Ex. 21.* Convert the surface integral

$$\oint (u \text{ grad } v - v \text{ grad } u) d\mathbf{S}$$

into a volume integral. In particular apply the result to  $v = 1/r$ . It should be remembered that the point  $O$  is to be excluded from the region of integration, on account of the discontinuity of  $1/r$ . Thus the surface integral must also extend over the small sphere enclosing  $O$  and having its normal directed toward this point.

### 13.\* Differential Vector Operations in Curvilinear Orthogonal Co-ordinates.

In many cases occurring in practice, it is more expedient to resolve a vector not in Cartesian co-ordinates, as done exclusively thus far, but to use other parameters which are often inherent in the problem itself. For example, in studying the propagation of electromagnetic waves over the earth's surface, spherical surface co-ordinates (latitude and longitude) are indicated. In space, a point is located by means of three parameters  $u, v, w$ . The rectangular co-ordinates are then functions of these parameters:

$$x = \phi_1(u, v, w), \quad y = \phi_2(u, v, w), \quad z = \phi_3(u, v, w). \quad . \quad (99)$$

If one co-ordinate, say  $u$ , is held constant, (99) will represent the parametric equation of the surface  $u = \text{const.}$  In the same way we get another surface for  $v = \text{const.}$ , and through every point in space there

\* Starred sections may be omitted without destroying the continuity.

will be three surfaces corresponding to the three co-ordinate planes. We restrict our considerations to those systems where the three co-ordinate surfaces form an orthogonal set, i.e. are normal to each other in pairs (fig. 26), as this type of curvilinear co-ordinate system is almost exclusively used. In the case of spherical space co-ordinates it is immediately evident that the three surfaces—sphere, radial cone and meridian plane—are mutually perpendicular. Elliptical co-ordinates are also orthogonal. Their surfaces consist of confocal ellipsoids of rotation, hyperboloids of either one or two sheets, and planes containing one or the other principal axis of the conics. If the co-ordinate *surfaces* are orthogonal, so will be the curves of intersection, called co-ordinate *lines*. The tangents drawn at an intersection of three

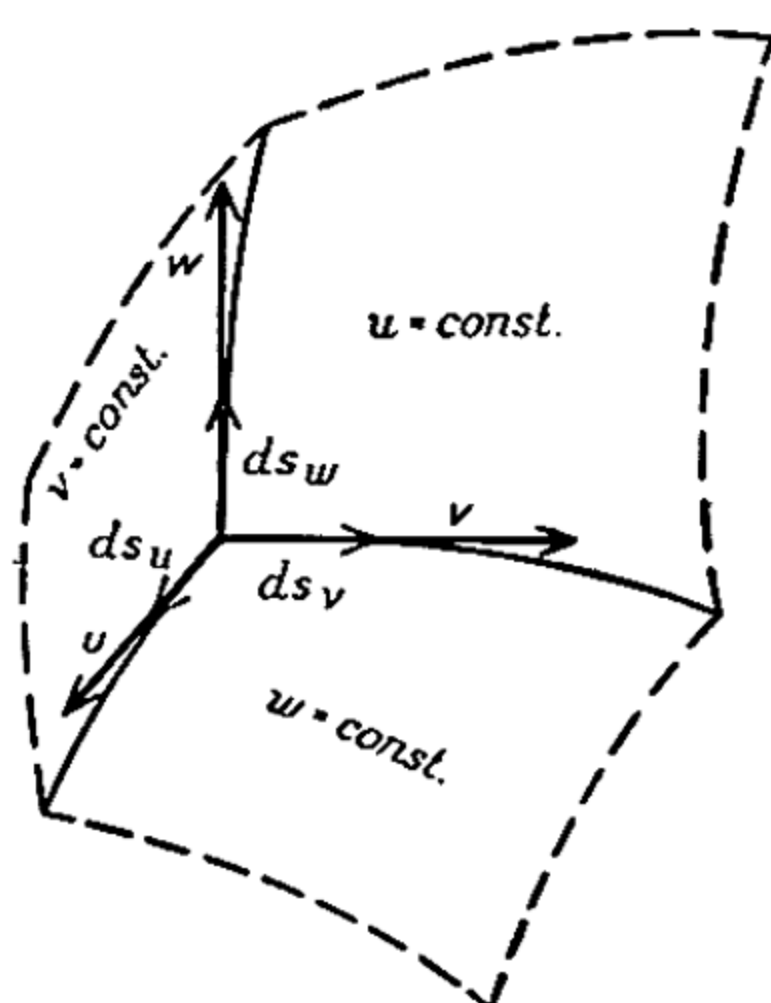


Fig. 26

surfaces will be characterized by the unit vectors  $u$ ,  $v$ ,  $w$ , corresponding to the Cartesian unit vectors  $i$ ,  $j$ ,  $k$ . There is an important difference between the two sets: the former change their direction as we go from point to point. If a vector  $A$  is given as a function of position, its respective components at the point  $P$  are understood to be its projections on the directions  $u$ ,  $v$ ,  $w$  drawn at the point in question. The scalar and vector products retain their form in this scheme and it is only when we pass to differential operations that the relations differ from their Cartesian forms. If the co-ordinate  $u$  is allowed to increase by  $du$ , there is a displacement in the  $u$  direction of amount

$$ds_u = U(u, v, w) du u, \quad . \quad . \quad . \quad (100)$$

where the function  $U(u, v, w)$  changes from one place to another. In an individual case its value may be read off from a figure. In the same way,

$$ds_v = V(u, v, w) dv v, \quad ds_w = W(u, v, w) dw w \quad . \quad (100a)$$



The components of the gradient of a scalar function  $\psi$  may be found at once as the projections in the  $u$ ,  $v$ ,  $w$  directions, and these projections represent the derivatives  $\partial\psi/\partial s$  in these directions. From (100),

$$\text{grad}_u \psi = \frac{1}{U} \frac{\partial \psi}{\partial u}, \quad \text{grad}_v \psi = \frac{1}{V} \frac{\partial \psi}{\partial v}, \quad \text{grad}_w \psi = \frac{1}{W} \frac{\partial \psi}{\partial w}. \quad (101)$$

In order to compute the divergence of a vector  $\mathbf{A}$  we start with the definition (40), p. 24, selecting as volume element a parallelepiped formed by co-ordinate surfaces. Consider the two sides normal to  $u$ . One of these is strictly perpendicular; the other is perpendicular as far as quantities of the second order. Only the  $u$ -component  $A_u$  will contribute to the surface integral. At the point  $u$  the area of the surface element is  $|d\mathbf{s}_v| |d\mathbf{s}_w| = (VW)_u dv dw$ , while at  $u + du$  it is  $(VW)_{u+du} dv dw$ , so that considering the opposite directions of the normals, the net contribution of both surfaces will be

$$-(A_u VW)_u dv dw + (A_u VW)_{u+du} dv dw = \frac{\partial(A_u VW)}{\partial u} du dv dw.$$

Forming the analogous expression for the other pairs of surfaces and dividing by the volume  $UVW du dv dw$ , we have

$$\text{div } \mathbf{A} = \frac{1}{UVW} \left[ \frac{\partial(A_u VW)}{\partial u} + \frac{\partial(A_v WU)}{\partial v} + \frac{\partial(A_w UV)}{\partial w} \right]. \quad (102)$$

Substituting the special value  $\text{grad } \psi$  for the vector  $\mathbf{A}$ , we have

$$\Delta\psi = \frac{1}{UVW} \left[ \frac{\partial \left( \frac{VW}{U} \frac{\partial \psi}{\partial u} \right)}{\partial u} + \frac{\partial \left( \frac{WU}{V} \frac{\partial \psi}{\partial v} \right)}{\partial v} + \frac{\partial \left( \frac{UV}{W} \frac{\partial \psi}{\partial w} \right)}{\partial w} \right]. \quad (103)$$

In computing the curl we proceed again as in the case of rectangular Cartesian co-ordinates by traversing, one after the other, three elementary rectangles lying, respectively, in the three co-ordinate planes. It must be remembered that the lengths of two opposite sides will differ somewhat. In the surface  $u = \text{const.}$  we get the contributions  $(A_v V)_w dv$  and  $(-A_v V)_{w+dw} dv$  by going along the  $v$  curves and the contributions  $-(A_w W)_v dw$  and  $(A_w W)_{v+dv} dw$  along the  $w$  curves. Expanding these expressions, adding them and dividing the result by



the area, we get the  $u$  component of the curl. The other two follow by permutation of the letters:

$$\left. \begin{aligned} (\text{curl } \mathbf{A})_u &= \frac{1}{VW} \left[ \frac{\partial(A_w W)}{\partial v} - \frac{\partial(A_v V)}{\partial w} \right], \\ (\text{curl } \mathbf{A})_v &= \frac{1}{WU} \left[ \frac{\partial(A_u U)}{\partial w} - \frac{\partial(A_w W)}{\partial u} \right], \\ (\text{curl } \mathbf{A})_w &= \frac{1}{UV} \left[ \frac{\partial(A_v V)}{\partial u} - \frac{\partial(A_u U)}{\partial v} \right]. \end{aligned} \right\} \quad (104)$$

The expression for the vector gradient in curvilinear co-ordinates is troublesome to derive, and since it is seldom needed it will not be deduced here.

*Ex. 22.* Compute  $\text{grad } \psi$ ,  $\text{div } \mathbf{A}$ ,  $\Delta \psi$  and  $\text{curl } \mathbf{A}$  in cylindrical co-ordinates  $(z, \rho, \phi)$  and in space polar co-ordinates  $(r, \phi, \theta)$ .

#### 14. Degeneration of the Vector Differential Operations at Surfaces of Discontinuity in the Field.

In proving the above relationships of vector analysis, we have hitherto imposed the condition that the point function be continuous in the region under consideration. However, the case often arises in physics where a scalar or vector point function has different values on two sides of a given surface. In this case, the limiting values defined by our differential operations become infinite. If, for example, we consider the divergence

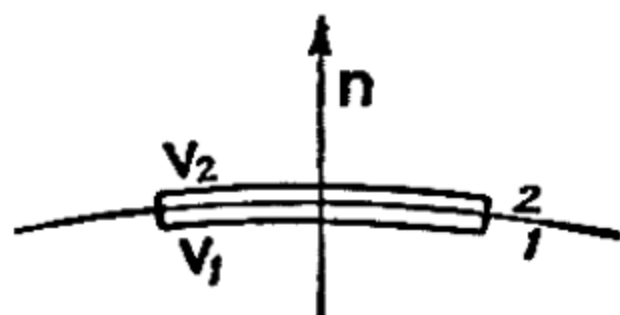


Fig. 27.

$$\text{div } \mathbf{v} = \lim_{\Delta \tau} \frac{1}{\Delta \tau} \oint \mathbf{v} \, d\mathbf{S},$$

this limit has a meaning only if the vector  $\mathbf{v}$  is continuous within the volume element  $\Delta \tau$ . Suppose, however, that a surface of discontinuity passes through  $\Delta \tau$ . Then we may choose as volume element a small cylinder whose bases  $\Delta \mathbf{S}$  are close to the surface of discontinuity, and whose lateral surface is perpendicular to it (fig. 27). With continuous point functions, the difference  $\mathbf{v}_2 - \mathbf{v}_1$  of the values on the two bases of the cylinder vanishes as the altitude of the cylinder is reduced to zero; here, however,  $\mathbf{v}_2 - \mathbf{v}_1$  becomes the difference of the values on the two sides of the surface of discontinuity. If we divide the integral  $\oint \mathbf{v} \, d\mathbf{S}$ , taken over the surface of the cylinder, by the volume  $\Delta \tau$ , the

result is infinite, for  $\Delta\tau$  vanishes to a higher order as the cylinder is made smaller. If we wish to obtain a limit in such cases, we must divide by the area of the base  $\Delta S$  rather than by the volume of the cylinder. The resulting limit is called the surface divergence ( $\text{Div } \mathbf{v}$ )

$$\text{Div } \mathbf{v} = \lim \frac{1}{\Delta S} \oint \mathbf{v} d\mathbf{S}. \quad . \quad . \quad . \quad (105)$$

If we introduce as unit normal vector a vector of length 1 pointing from side 1 toward side 2, then, with a sufficiently small surface element, we have for the lower side

$$\int_1 \mathbf{v} d\mathbf{S} = -n v_1 \Delta S,$$

and for the upper side

$$\int_2 \mathbf{v} d\mathbf{S} = +n v_2 \Delta S.$$

The negative sign is to be used on the lower side, since the normals are to point outward in forming the surface integral, but the vector  $\mathbf{n}$  was taken to point inward. The contribution of the lateral cylindrical surface can be made an infinitesimal of higher order by reducing the altitude of the cylinder. We thus have, as a limit, the contributions of the bases of the cylinder, and so

$$\text{Div } \mathbf{v} = \lim \frac{1}{\Delta S} n(v_2 - v_1) \Delta S = n(v_2 - v_1). \quad . \quad (106)$$

Since the definition by means of the limit of a surface integral divided by the enclosed volume is not confined to the divergence, but represents the general co-ordinate-free definition of the  $\nabla$ -operator, we obtain for surfaces of discontinuity, in analogy with the ordinary  $\nabla$ -operator, a reduced operator—the “surface nabla”—denoted by the sign  $\|$ . Its particular meaning depends upon whether it is applied to a scalar- or a vector-point function, and will be made clear by the following:

$$\text{Surface gradient: } \| u = \text{Grad } u = n(u_2 - u_1). \quad . \quad . \quad . \quad (107)$$

$$\text{Surface divergence}_x \| \mathbf{v} = \text{Div } \mathbf{v} = n(v_2 - v_1). \quad . \quad . \quad . \quad (108)$$

$$\text{Surface curl: } [\| \mathbf{v}] = \text{Curl } \mathbf{v} = [n(v_2 - v_1)]. \quad . \quad . \quad . \quad (109)$$

$$\text{Surface vector gradient: } \| \cdot \mathbf{v} = \text{Grad} \cdot \mathbf{v} = n \cdot (v_2 - v_1). \quad . \quad (110)$$

If a surface of discontinuity passes through the region of integration  $V$ , the theorem of Gauss must be supplemented by a term corresponding to the surface divergence at this surface of discontinuity. This is done as follows: A closed surface is drawn surrounding the surface of discontinuity closely (fig. 28). This separates the latter surface

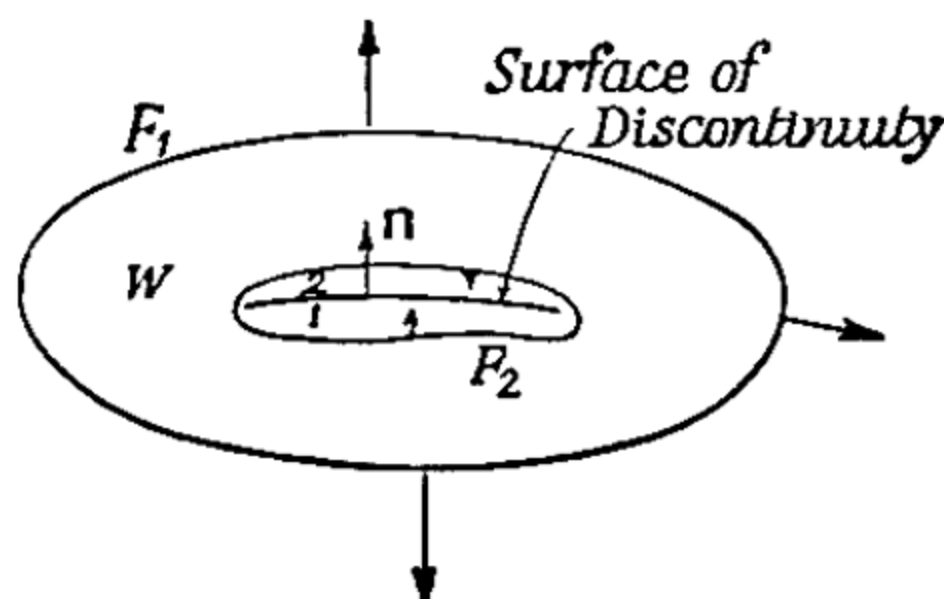


Fig. 28

from the region of integration. Gauss's Theorem is now applied to the remaining volume, remembering that the new surface  $F_2$ , as well as the outer surface  $F_1$ , contributes to the surface integral:

$$\oint_{F_1} \mathbf{v} d\mathbf{S} + \oint_{F_2} \mathbf{v} d\mathbf{S} = \int_V \text{div } \mathbf{v} d\tau.$$

If  $\mathbf{n}$  is again used to indicate the unit normal vector of the surface of discontinuity, pointing from side 1 to side 2, then

$$\oint_{F_2} \mathbf{v} d\mathbf{S} = \int \mathbf{n} v_1 dS - \int \mathbf{n} v_2 dS = - \int \text{Div } \mathbf{v} dS,$$

whence 
$$\oint_{F_1} \mathbf{v} d\mathbf{S} = \int_V \text{div } \mathbf{v} d\tau + \int \text{Div } \mathbf{v} dS. \quad . \quad . \quad (111)$$

How is the reduced nabla operation to be performed on a product of point functions which are discontinuous on a given surface? If the calculational rules of § 12 (p. 38) are formally applied here, what meaning is to be assigned to a quantity with subscript  $c$ ? A simple manipulation makes this clear. We have

$$\mathbf{n}(u_2 v_2 - u_1 v_1) = \mathbf{n} \left[ \frac{v_1 + v_2}{2} (u_2 - u_1) + \frac{u_1 + u_2}{2} (v_2 - v_1) \right].$$

If we designate the average values  $(v_1 + v_2)/2$  in the customary way, by a horizontal bar over the symbols, then

$$\text{Grad } uv = \parallel uv_c + \parallel u_c v = \mathbf{n} \bar{v}_1 (u_2 - u_1) + \mathbf{n} \bar{u}_1 (v_2 - v_1). \quad (112)$$



The subscript  $c$  on a quantity standing after the surface nabla thus signifies that this quantity is to be replaced by the mean of the values on the two sides of the surface of discontinuity. In this way all the calculational rules of § 12 (p. 38) may be applied formally to the degenerate differential operations.

### 15.\* Fundamentals of the Matrix Calculus.\*

On p. 33 the coefficients  $a_{ik}$  fulfilling certain transformation conditions were designated as tensor components. It proves useful in many branches of physics to introduce quite general arrays of such coefficients as mathematical quantities. An array of this kind need not be square; it may have  $m$  rows and  $n$  columns, and would then be called a *matrix* of order  $m, n$ . The individual coefficients  $a_{ik}$  are called elements of the matrix. The entire matrix is designated by a bold-face letter: **A**.

Two matrices are equal when and only when their corresponding elements are identical. Thus a matrix equation replaces  $mn$  numerical equations. A matrix having all its elements equal to zero is called a null matrix, **O**. A square matrix having only its diagonal elements different from zero is called a diagonal matrix. If all these diagonal elements are equal to unity the matrix is called the unit matrix, **E**.

A square matrix is said to be *symmetric* if each element is equal to the one which is in a symmetric position to it with respect to the diagonal, that is, if  $a_{ik} = a_{ki}$ . A square matrix is said to be *anti-symmetric* (or skew symmetric) if each element is opposite in sign to its "reflection" in the diagonal, that is, if  $a_{ik} = -a_{ki}$  and further, if the diagonal elements are zero ( $a_{ii} = 0$ ). The matrix **A** formed by interchanging rows and columns of **A** is called the *transpose* of **A**. A symmetric matrix is, then, its own transpose.

As in the case of tensors, which are merely a special kind of matrix, the sum of two matrices is defined as the matrix in which each element is the sum of the corresponding elements of the separate matrices. Logically, the product of a matrix by a number  $\mu$  is then taken to be the matrix, each of whose elements is  $\mu$  times that of the original matrix.

Up to this point we have given merely an alternative way of writing the results of p. 33. The matrix calculus becomes really fruitful when we define the product of two matrices. This permits us to replace long-winded summation formulas by a new symbolism. We start again with a linear transformation which, through the matrix **A**, gives us the variables  $z_k$  in terms of the variables  $y_k$ . If, in turn, the variables  $y_k$  are given in terms of  $x_k$  through the medium of a matrix **B**, then the

\* Although the matrix calculus will be used very little in the text, a brief account of the basic principles is given here because of the increasing use of this branch of mathematics.



$z_k$  are themselves linear functions of the  $x_k$ . It is readily computed that the coefficients  $c_{ik}$  are given by

$$c_{ik} = \sum_l a_{il} b_{lk} \quad . \quad . \quad . \quad . \quad . \quad . \quad (113)$$

This means that the matrix element  $c_{ik}$  is obtained by multiplying each element of the  $i$ th row of **A** by the corresponding element of the  $k$ th column of **B** and adding the products. It is logical to designate this combination as the product of the two matrices, since it can be shown readily that it has the property of distributivity

$$\mathbf{A}(\mathbf{B} + \mathbf{C}) = \mathbf{AB} + \mathbf{AC} \quad . \quad . \quad . \quad . \quad (114)$$

as well as of associativity

$$\mathbf{A}(\mathbf{BC}) = (\mathbf{AB})\mathbf{C}. \quad . \quad . \quad . \quad . \quad . \quad (115)$$

On the other hand, like the vector product, it does not possess commutativity: The result is quite different from the previous if the  $z_k$  and the  $y_k$  are related by the matrix **B**, and the  $y_k$  and the  $x_k$  by the matrix **A**.

It may happen that the product of two matrices is independent of their order of multiplication; the two are then said to be *commutative*; for example, the unit matrix **E** is commutative with every matrix. The product of any matrix with the null matrix **Q** is always the null matrix itself. On the other hand, the product of two matrices, neither of which is the null matrix, may happen to be the null matrix; it is not proper to conclude that if the matrix product vanishes one of the factors must be the null matrix. In view of the rule for forming the elements of the product matrix, we have

$$\widetilde{\mathbf{AB}} = \widetilde{\mathbf{B}} \widetilde{\mathbf{A}} \quad . \quad . \quad . \quad . \quad . \quad . \quad (116)$$

The definition of matrix product is not confined to square matrices. The procedure designated by (113) can be carried out as long as the number of columns of the first matrix is equal to the number of rows of the second. Thus matrices of order  $m, n$  may always be multiplied by those of order  $n, r$ ; the product will be a matrix of order  $m, r$ . This permits us to write vectors as single-column or single-row matrices, e.g.

$$\boldsymbol{r} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} \quad . \quad . \quad . \quad . \quad . \quad . \quad (117)$$

The transformation prescribed by a matrix **A** can be written

$$\boldsymbol{r}' = \begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \mathbf{A}\boldsymbol{r}. \quad . \quad . \quad . \quad . \quad . \quad (118)$$

The product of a matrix  $\mathbf{A}$  of order 3,3 with the matrix  $\mathbf{r}$  of order 3,1 is a matrix of order 3,1 and so is again a vector.

If the determinant formed from the coefficients of a given square matrix is different from zero, then the matrix which represents  $y_k$  as functions of  $x_k$  may be solved for the  $x_k$  as functions of  $y_k$ . The new scheme which results is the inverse, or *reciprocal*, of the original matrix and is designated  $\mathbf{A}^{-1}$ . Since application of  $\mathbf{A}$  to the  $x_k$  gives us the  $y_k$ , and application of  $\mathbf{A}^{-1}$  to the  $y_k$  again gives us the  $x_k$ , it must be true that

$$\mathbf{A}^{-1}\mathbf{A} = \mathbf{A}\mathbf{A}^{-1} = \mathbf{E}. \quad . \quad . \quad . \quad . \quad . \quad (119)$$

The second equality may be verified by starting with the  $y_k$ . Since the inverse of the inverse matrix must again be the original matrix,

$$(\mathbf{A}^{-1})^{-1} = \mathbf{A}. \quad . \quad . \quad . \quad . \quad . \quad (120)$$

Moreover,

$$(\mathbf{AB})^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}. \quad . \quad . \quad . \quad . \quad . \quad (121)$$

This is true because if we multiply both sides of the equation by  $\mathbf{AB}$ ,

$$\mathbf{E} = \mathbf{ABB}^{-1}\mathbf{A}^{-1} = \mathbf{AA}^{-1} = \mathbf{E}, \quad . \quad . \quad . \quad . \quad . \quad (122)$$

on account of the associative law.

An *orthogonal* matrix is one in which the sum of the products of corresponding members of two rows has the following properties:

$$\left. \begin{aligned} \sum_k a_{ik}a_{lk} &= 0 \quad \text{for } l \neq i \\ &= 1 \quad \text{for } l = i \end{aligned} \right\}. \quad . \quad . \quad . \quad . \quad . \quad (123)$$

The definition may be expressed in another way. The product sums defined above represent the product of the matrix  $\mathbf{A}$  with its transpose  $\widetilde{\mathbf{A}}$ . If a matrix is orthogonal, then

$$\mathbf{A} \widetilde{\mathbf{A}} = \mathbf{E}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (124)$$

However, this is also the definition of the matrix  $\mathbf{A}^{-1}$  which is reciprocal to  $\mathbf{A}$ . Thus an orthogonal matrix has the property that its transpose is the same as its reciprocal. An example of an orthogonal matrix is the set of direction cosines of two rectangular co-ordinate systems that differ in orientation (equation (64), p. 33).

The matrix calculus may be extended to include matrices with complex elements. The transposed matrix is then one where, in addition to the interchange of rows and columns, each element is replaced by its complex conjugate. A symmetric matrix then corresponds with the so-called Hermitian matrix, and the orthogonal matrix, because of  $\mathbf{U} \widetilde{\mathbf{U}} = \mathbf{E}$  becomes identical with the unit matrix.

*Ex. 22a.* The question posed in *Ex. 18* (p. 37) may be formulated in matrix notation as follows: Given an  $n$ -dimensional symmetric square matrix  $\mathbf{A}$ . Find the vectors  $\mathbf{r}$  (single-column matrices) for which

$$\mathbf{A}\mathbf{r} = \lambda\mathbf{r} \quad \text{or} \quad (\mathbf{A} - \lambda\mathbf{E})\mathbf{r} = 0.$$

Show that (a) this leads to an equation of  $n$ th order in  $\lambda$ ; (b) the  $\mathbf{r}$  corresponding to various values of  $\lambda$  are mutually perpendicular (since it is only the *ratios* of the components of  $\mathbf{r}$  that enter, their magnitude may be taken equal to unity); (c) the matrix formed by using these vector components as rows is an orthogonal matrix,  $\mathbf{S}$ ; (d) the matrix  $\mathbf{S}^{-1}\mathbf{A}\mathbf{S}$  is a diagonal matrix with the roots  $\lambda_i$  occupying the diagonal.

## CHAPTER II

### MATHEMATICAL REPRESENTATION OF PERIODIC PHENOMENA; THEORY OF VIBRATIONS AND WAVES

#### 1. Simple Harmonic Vibrations.

The simplest temporally periodic process (vibration) is described mathematically by the sine or cosine function. If the process repeats itself  $\nu$  times per second, the function is either

$$u = A \sin 2\pi \nu t \quad \text{or} \quad u = A \cos 2\pi \nu t. \quad . . . (1)$$

$A$  is called the *amplitude* and  $\nu$  the *frequency* of the vibration. Besides the frequency  $\nu$ , we introduce the *angular frequency*  $\omega = 2\pi\nu$  to simplify the notation. A phenomenon described by a simple sine or cosine function is not only the simplest vibration mathematically, but physically as well. It is called a *simple harmonic vibration*.

The computations are considerably simplified by using imaginary exponentials instead of the trigonometric functions with their cumbersome addition theorems. The connexion between the two types of functions is given by the *Euler Formula*

$$e^{i\omega t} = \cos \omega t + i \sin \omega t. \quad . . . . . (2)$$

This notation also leads to an extremely important representation of a vibration, using the complex plane (cf. Chap. III); for

$$z = Ae^{i\omega t} \quad . . . . . (3)$$

represents a complex number whose representative point, or *index*, as we shall call it, describes a circle of radius  $A$  with angular velocity  $\omega$ . The projections on the real and imaginary axes are respectively

$$x = R(z) = A \cos \omega t; \quad y = I(z) = A \sin \omega t. \quad . . (4)$$

But since physics deals with real quantities, the final results of a computation using complex numbers must admit of translation into real magnitudes. This is very simple; an equation involving complex numbers means that the real as well as the imaginary parts satisfy the equation. We can therefore take either the real or the imaginary part of the equation as the physical statement.

The square of the amplitude is often of importance. It is most



readily obtained from the complex representation by multiplying the vibration function by its conjugate:

$$A^2 = z\bar{z}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

The great advantage of using complex numbers manifests itself in the case where two vibrations of the same frequency, but of different phase, are to be combined. If there is but one vibration function, the point at which we begin to count the time is of no consequence, and so we write this function in the form (1) or (3). If, however, we are required to add a second vibration to this one, the new oscillation will not, in general, attain its maximum value at the same instant as the original one, but at a certain interval before or after. In this case we say that there is a *difference of phase*  $\delta$  between the two vibrations. In the real representation, then,

$$u_1 = A_1 \cos \omega t, \quad u_2 = A_2 \cos(\omega t + \delta),$$

and the resultant of the two oscillations is

$$u_1 + u_2 = A_1 \cos \omega t + A_2 \cos \delta \cos \omega t - A_2 \sin \delta \sin \omega t. \quad (6)$$

The right member may be written as a single cosine vibration by transforming the coefficients of  $\sin \omega t$  and  $\cos \omega t$  so that they assume the character of the sine and cosine of an angle  $\phi$ , multiplied by the same coefficient:

$$\begin{aligned} & (A_1 + A_2 \cos \delta) \cos \omega t - A_2 \sin \delta \sin \omega t \\ &= \sqrt{(A_1 + A_2 \cos \delta)^2 + A_2^2 \sin^2 \delta} \left\{ \frac{A_1 + A_2 \cos \delta}{\sqrt{(A_1 + A_2 \cos \delta)^2 + A_2^2 \sin^2 \delta}} \cos \omega t \right. \\ & \quad \left. - \frac{A_2 \sin \delta}{\sqrt{(A_1 + A_2 \cos \delta)^2 + A_2^2 \sin^2 \delta}} \sin \omega t \right\}. \quad \cdot \cdot \cdot \cdot \cdot (7) \end{aligned}$$

Put  $\frac{A_1 + A_2 \cos \delta}{\sqrt{(A_1 + A_2 \cos \delta)^2 + A_2^2 \sin^2 \delta}} = \cos \phi; \quad \dots (8)$

$$\frac{A_2 \sin \delta}{\sqrt{(A_1 + A_2 \cos \delta)^2 + A_2^2 \sin^2 \delta}} = \sin \phi. \quad \cdot \cdot \quad (9)$$

Then 
$$u_1 + u_2 = \sqrt{A_1^2 + A_2^2 + 2A_1A_2 \cos \delta} (\cos \omega t + \phi) = A \cos(\omega t + \phi). \quad (10)$$

According to the expression under the radical, the amplitude  $A$  of the resulting vibration is given as the third side of the triangle having the amplitudes  $A_1$  and  $A_2$  as adjacent sides, and  $\delta$  as exterior angle (fig. 1). The difference in phase  $\phi$  between the resultant vibration and  $u_1$  is the angle between the sides  $A$  and  $u_1$  of this triangle.

This construction corresponds exactly with the addition, in the

Gauss plane, of two complex numbers, which add like vectors. To obtain the resultant of two vibrations having equal frequencies, but differing in phase, add the corresponding complex numbers. The amplitude and phase of the resultant are given, respectively, by the length and direction of the radius vector representing this sum.

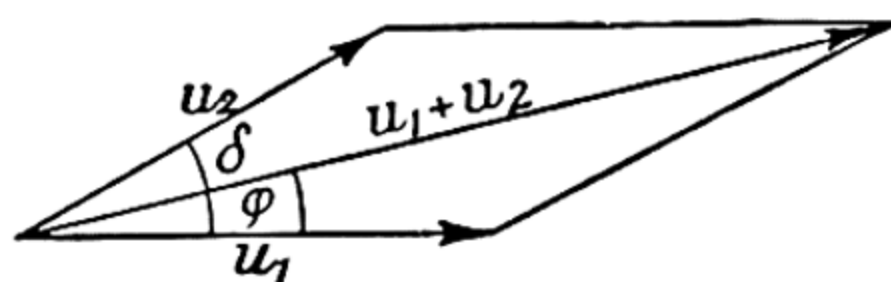


Fig. 1

In electrical engineering, where this method is most frequently used, this representation is referred to as the “vector diagram”, on account of the way in which the quantities are combined. It must be remembered, however, that the complex numbers by means of which the vibration phenomenon is represented are not vectors in the physical sense; the latter are *spatial*, and are not confined to a plane. For this reason, it would be more correct to speak of the “index diagram” than of the vector diagram. This diagram may be drawn for any instant of time; but since the phase difference is constant, the entire triangle rotates as a rigid figure with the angular speed  $\omega$ , as the time advances. We may therefore choose any position—for example, the real axis—for the line of the first vibration.

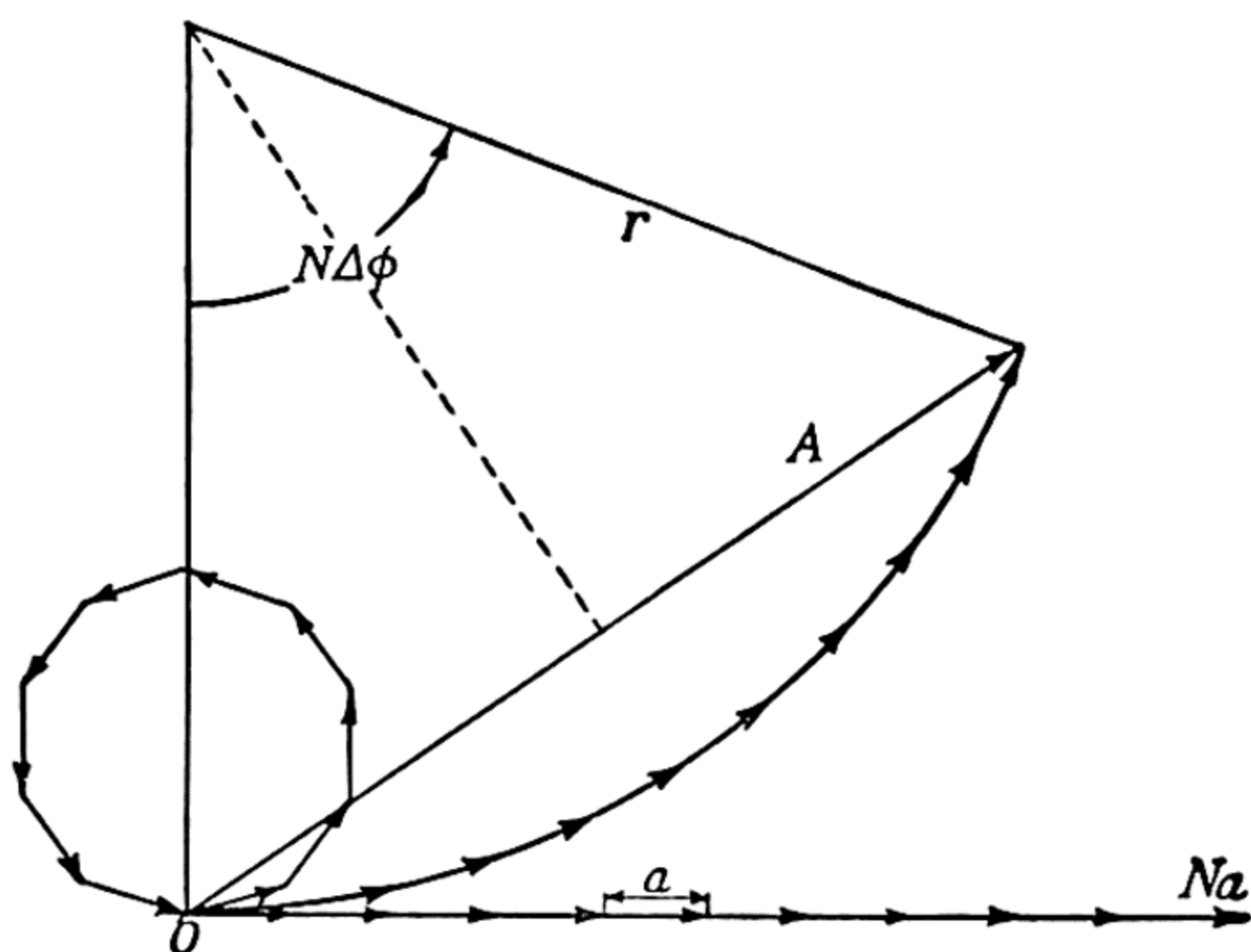


Fig. 2

If more than two vibrations are to be compounded, we obtain the index of the resultant as before by laying off the individual indices in succession—like vectors—connecting the initial point of the first index with the terminus of the last. A case occurring frequently in practice is the combination of a large number of vibrations of equal amplitude, there being a constant difference of phase  $\Delta\phi$  between successive vibrations. The vector diagram then becomes a portion of a polygon, which approaches a circular arc as the individual amplitudes

decrease (fig. 2). The phase difference between the first and the  $N$ th vibration is  $(N - 1)\Delta\phi$ , which may be written  $N\Delta\phi$  if  $N$  is large. The greatest value of the resultant amplitude is  $Na$ , obtained when  $\Delta\phi = 0$ . As  $\Delta\phi$  increases, the broken line becomes a circular arc of greater and greater curvature. If the arc closes to form a full circle, i.e. if  $\Delta\phi = 2\pi/N$ , the resulting amplitude vanishes. As the phase difference further increases, a small arc remains after the circle is closed. This arc again yields a finite value of the resultant, until, at  $\Delta\phi = 4\pi/N$ , a second zero point is reached, and so on.

For arbitrary phase differences  $\Delta\phi$  we obtain, from fig. 2, a resultant amplitude

$$A = 2r \sin \frac{N\Delta\phi}{2},$$

$$\text{where } rN\Delta\phi = Na, \text{ i.e. } A = Na \frac{\sin \frac{N\Delta\phi}{2}}{\frac{N\Delta\phi}{2}}, \quad \dots \dots \dots (11)$$

$$\text{or, putting } N\Delta\phi = \phi, \quad A = Na \frac{\sin \frac{\phi}{2}}{\frac{\phi}{2}}. \quad \dots \dots \dots (11')$$

The representation by complex numbers also gives a simple construction for the derivative. If

$$z = Ae^{i\omega t},$$

then

$$\frac{dz}{dt} = i\omega Ae^{i\omega t} = i\omega z.$$

This is a complex number obtained by multiplying  $z$  by  $\omega$  and turning it through an angle  $\pi/2$  in the counter-clockwise direction. Differentiating again,

$$\frac{d^2z}{dt^2} = -\omega^2 Ae^{i\omega t} = -\omega^2 z.$$

This is a complex number whose index again falls in the line  $Oz$ , but in the opposite direction. We see, in addition, that a simple harmonic vibration satisfies the differential equation

$$\frac{d^2z}{dt^2} + \omega^2 z = 0, \quad \dots \dots \dots (12)$$

which is called the differential equation of an (undamped) simple harmonic vibration. Similarly, the indefinite integral

$$\int z dt = \frac{1}{i\omega} Ae^{i\omega t} = -\frac{i}{\omega} z$$



is found by rotating the index  $z$  through an angle  $\pi/2$  in the clockwise direction, and dividing by  $\omega$ .

## 2. Representation of more complicated Periodic Phenomena by Series of Harmonic Terms. Fourier Series. The Fourier Integral.

Let a purely arbitrary process, which is repeated  $\nu$  times per second, be represented by  $u = f(t)$ . That is to say, we have

$$f\left(t + \frac{1}{\nu}\right) = f(t). \quad . \quad . \quad . \quad . \quad . \quad (13)$$

Let us make the proviso that  $f(t)$  is not to pass through infinitely many maxima and minima within the time of one vibration, and that the integral  $\int f(t) dt$  is to converge even if the function has a finite number of infinite values in this range. A finite number of discontinuities is also permissible. Under these conditions the function  $f(t)$  may be represented over a complete period, and thus from  $t = -\infty$  to  $t = +\infty$ , except at the discontinuities, by a series of simple harmonic functions, the frequencies in which are integral multiples of the fundamental frequency. Such series are called *Fourier Series*, after their discoverer. For a proof of the possibility of developing  $f(t)$  in a Fourier series under these very general conditions, we must refer to textbooks on mathematics. Here we shall give only the method of determining the coefficients. It is most convenient to start from the complex representation and write

$$\begin{aligned} f(t) &= a_0 + a_1 e^{i\omega t} + a_2 e^{2i\omega t} + \dots + a_n e^{ni\omega t} + \dots \\ &\quad + a_{-1} e^{-i\omega t} + a_{-2} e^{-2i\omega t} + \dots + a_{-n} e^{-ni\omega t} + \dots \\ &= \sum_{n=-\infty}^{n=+\infty} a_n e^{ni\omega t}. \quad . \quad . \quad . \quad (14) \end{aligned}$$

Since the left member is real, the coefficients of the series on the right must be such that no imaginary terms occur. To determine  $a_0$ , we integrate both sides over one complete period, i.e. from 0 to  $2\pi/\omega$ . Because of the periodicity of the complex exponential function, all contributions from the terms on the right vanish, with the exception of  $a_0$ , for

$$\int_0^{2\pi/\omega} a_n e^{ni\omega t} dt = \frac{a_n}{ni\omega} \left[ e^{ni\omega t} \right]_0^{2\pi/\omega} = \frac{a_n}{ni\omega} (e^{2\pi ni} - e^0) = 0. \quad (15)$$

On the other hand,  $a_0$  is determined by the equation

$$\int_0^{2\pi/\omega} f(t) dt = \frac{2\pi}{\omega} a_0, \quad \text{or} \quad a_0 = \frac{\omega}{2\pi} \int_0^{2\pi/\omega} f(t) dt = \overline{f(t)}.^* \quad (16)$$

\*  $\overline{f(t)}$  is the usual notation for the mean value of  $f(t)$ .



To determine the other coefficients, we multiply both sides by  $e^{-ni\omega t}$  and integrate, as before, from 0 to  $2\pi/\omega$ . Again, all terms on the right cancel out (because of the periodicity of the imaginary exponentials) except the  $a_n$ -term, which contains no exponential factor, and which gives the value  $(2\pi/\omega)a_n$  on integration. Thus we have

$$a_n = \frac{\omega}{2\pi} \int_0^{2\pi/\omega} f(t) e^{-ni\omega t} dt, \quad . . . . . (17)$$

and similarly 
$$a_{-n} = \frac{\omega}{2\pi} \int_0^{2\pi/\omega} f(t) e^{ni\omega t} dt. \quad . . . . . (18)$$

The reduction to the usual real form of the Fourier series,

$$\begin{aligned} f(t) &= b_0 + b_1 \cos \omega t + \dots b_n \cos n\omega t + \dots c_1 \sin \omega t + \dots c_n \sin n\omega t + \dots \\ &= \sum_0^{\infty} b_n \cos n\omega t + \sum_1^{\infty} c_n \sin n\omega t, \quad . . . . . (19) \end{aligned}$$

is very simple. If the real and imaginary parts of each term in the complex form are separated, and the terms in  $a_n$  and  $a_{-n}$  are paired, then we have

$$\left. \begin{aligned} a_n e^{ni\omega t} &= \frac{\omega}{2\pi} (\cos n\omega t + i \sin n\omega t) \\ &\quad \left\{ \int_0^{2\pi/\omega} f(t) \cos n\omega t dt - i \int_0^{2\pi/\omega} f(t) \sin n\omega t dt \right\}; \\ a_{-n} e^{-ni\omega t} &= \frac{\omega}{2\pi} (\cos n\omega t - i \sin n\omega t) \\ &\quad \left\{ \int_0^{2\pi/\omega} f(t) \cos n\omega t dt + i \int_0^{2\pi/\omega} f(t) \sin n\omega t dt \right\}; \\ a_n e^{ni\omega t} + a_{-n} e^{-ni\omega t} &= \frac{\omega}{\pi} \cos n\omega t \int_0^{2\pi/\omega} f(t) \cos n\omega t dt \\ &\quad + \frac{\omega}{\pi} \sin n\omega t \int_0^{2\pi/\omega} f(t) \sin n\omega t dt; \\ \text{i.e. } b_0 = \overline{f(t)}; \quad b_n &= \frac{\omega}{\pi} \int_0^{2\pi/\omega} f(t) \cos n\omega t dt; \quad c_n = \frac{\omega}{\pi} \int_0^{2\pi/\omega} f(t) \sin n\omega t dt. \end{aligned} \right\} (20)$$

It is to be noted that the summation extends from  $n = -\infty$  to  $n = +\infty$  in the complex representation, but from 0 or from 1 to  $+\infty$  if the real representation is used.\*

By inserting the expressions for the coefficients, we can represent the Fourier Series by a single summation. However, it is necessary to

\* In most physical problems  $f(t)$  is not given by an analytical expression, but by an empirical curve, so that the integrals must be evaluated numerically or graphically. Special computing methods as well as mechanical devices called *harmonic analysers* have been worked out for this purpose.

assign a new symbol for the variable of integration, since otherwise  $t$  would be used for two different purposes. If the integration variable is called  $\alpha$ , then

$$\begin{aligned} f(t) &= \sum_{n=-\infty}^{+\infty} \frac{\omega}{2\pi} e^{ni\omega t} \int_0^{2\pi/\omega} f(\alpha) e^{-ni\omega\alpha} d\alpha \\ &= \sum_{n=-\infty}^{+\infty} \frac{\omega}{2\pi} \int_0^{2\pi/\omega} f(\alpha) e^{ni\omega(t-\alpha)} d\alpha. \quad . \quad . \quad . \quad (21) \end{aligned}$$

We introduce the time of one complete vibration—the period  $T = 2\pi/\omega$ . Also, on account of the periodicity of the function  $f(t)$ , we can take the range of integration from  $-T/2$  to  $+T/2$  instead of from 0 to  $T$ . Then there results

$$f(t) = \sum_{n=-\infty}^{+\infty} \frac{1}{T} \int_{-T/2}^{+T/2} f(\alpha) e^{\frac{2\pi ni}{T}(t-\alpha)} d\alpha. \quad . \quad . \quad (22)$$

At this point we can pass to non-periodic functions, which may be looked upon as extreme cases of periodic functions whose period extends from  $-\infty$  to  $+\infty$ . If we put  $1/T = \Delta s$ , then

$$f(t) = \sum_{n=-\infty}^{+\infty} \Delta s \int_{-T/2}^{+T/2} f(\alpha) e^{2\pi ni(t-\alpha)\Delta s} d\alpha. \quad . \quad . \quad (23)$$

Now the definite integral  $\int_0^\infty \phi(s) ds$  is defined as the limit, for  $\Delta s$  infinitely small, of the sum

$$\sum_{n=0}^{n=\infty} \phi(n\Delta s) \Delta s.$$

$$\text{Also, } \int_{-\infty}^{+\infty} \phi(s) ds = \int_{-\infty}^0 \phi(s) ds + \int_0^\infty \phi(s) ds = \sum_{-\infty}^{+\infty} \phi(n\Delta s) \Delta s. \quad (24)$$

From this result it follows that, as  $T$  grows beyond all bound, the expression (23) passes over into the *Fourier Integral*

$$f(t) = \int_{-\infty}^{+\infty} ds \int_{-\infty}^{+\infty} f(\alpha) e^{2\pi is(t-\alpha)} d\alpha = \int_{-\infty}^{+\infty} e^{2\pi ist} ds \int_{-\infty}^{+\infty} f(\alpha) e^{-2\pi is\alpha} d\alpha. \quad (25)$$

The second form discloses the meaning of this identity: the non-periodic function  $f$  may be expressed by a continuous series of harmonic periodic functions whose amplitudes are given as functions of the sequence number  $s$  by

$$\int_{-\infty}^{+\infty} f(\alpha) e^{-2\pi is\alpha} d\alpha.$$

The possibility of such a representation is of great importance in the analytical treatment of functions which are otherwise not expressible

by a unified mathematical expression. Let there be given, for example, a simple harmonic vibration which, however, is not of indefinite duration, i.e. we suppose it to start at  $t = 0$  and cease at  $t = T$ . Such a vibration cannot be represented by  $e^{2\pi i \nu t}$ , since this is the expression for a vibration existing from  $t = -\infty$  to  $t = +\infty$ . We can, nevertheless, insert the function  $f$ , so defined, in the Fourier Integral and determine the distribution of frequencies therein. This distribution has, naturally, a maximum for  $s = \nu$ . The shorter the time the vibration lasts, the wider the "frequency band" covered by it.

By taking the quantities in pairs and applying the Euler Formula, as in the case of the Fourier Series (p. 55), we can reduce the complex form of the Fourier Integral to a real form:

$$f(t) = 2 \int_0^\infty ds \int_{-\infty}^{+\infty} f(a) \cos 2\pi s(t - a) da. \quad . \quad . \quad (26)$$

*Ex. 23.* Develop, in a Fourier Series, the periodic function which has the value 0 from 0 to  $T/2$ , and the value 1 from  $T/2$  to  $T$ , then 0 once more, and so on—sometimes called a "square wave".

*Ex. 24.* Express  $e^{-at} \cos \omega t$  by a continuous series of undamped harmonic vibrations, and find the frequency of the component vibration for which the value of the square of the amplitude is half the maximum value.

### 3. Modulated Vibrations and Beats.

In wireless telephony we deal with "modulated" vibrations in the transmitter. These are oscillations in which the peak amplitude itself is a periodic function of the time. The amplitude, however, changes slowly compared with the frequency of the actual vibration. The latter vibration—the so-called "carrier wave"—has a frequency of the order of  $1,000,000 \text{ sec}^{-1}$ , while the frequency of modulation—which is the frequency of the radiated tone—is around  $500 \text{ sec}^{-1}$ . Using real functions, a modulated frequency of this kind may be represented by

$$u = A \cos \omega_1 t \cos \omega_2 t, \quad (\omega_2 \gg \omega_1). \quad . \quad . \quad . \quad (27)$$

But, by a familiar formula of trigonometry,

$$2 \cos \omega_1 t \cos \omega_2 t = \cos(\omega_2 + \omega_1)t + \cos(\omega_2 - \omega_1)t. \quad . \quad (28)$$

The use of this formula enables us to write (27) in the form

$$u = \frac{A}{2} \cos(\omega_2 + \omega_1)t + \frac{A}{2} \cos(\omega_2 - \omega_1)t. \quad . \quad . \quad (29)$$

This signifies the combination of two vibrations of equal amplitudes, and having frequencies  $\omega_2 + \omega_1$  and  $\omega_2 - \omega_1$ , which are very close together. In practice, this type of vibration may be produced in either of these two ways, i.e. by modulating a carrier frequency, or by combining two vibrations which differ but little in frequency. The



latter case, because of its well-known acoustical manifestation, is known as the phenomenon of *beats*. The angular frequency of the beats is  $\omega_2 - \omega_1$ . It is to be remembered, however, that the phenomenon itself is correctly represented by either (27) or (29).

The occurrence of beats represents the simplest case of an *almost-periodic* phenomenon. If, for example,  $\omega_1$  and  $\omega_2$  are not commensurable, then, strictly speaking, the function  $u$  has no definite period, i.e. there is no fixed interval  $T$  such that the value of  $u$  at any time  $t$  is repeated at time  $t + T$ . Nevertheless, within a certain latitude of departure  $\epsilon$ , a "time of recurrence" may be defined.

#### 4. Combination of Vibrations along different Axes. Lissajous' Figures.

Thus far we have considered perfectly general periodic functions  $u$ , which could be represented either in a co-ordinate system with  $t$  as abscissa and  $u$  as ordinate, or in the Argand diagram of the complex plane. Now we wish to give our function a particular meaning: let  $u$  be the co-ordinate  $x$  of a point moving along the  $x$ -axis. If, now, the other co-ordinates (components of the radius vector) are also periodic functions of the time, we inquire the nature of the curve described by  $P$ , the extremity of the radius vector. We restrict ourselves to the important case of motion in a plane, the  $z$ -co-ordinate being omitted. The motion in the  $xy$ -plane is to be distinguished, of course, from rotation of the index in a vector diagram. If we desire to employ the latter representation here also, we must introduce a complex plane with rotating index for *each* co-ordinate. For the discussion of the orbit, we thus find it simpler to start from the real representation and write

$$x = a \cos \omega t, \quad y = b \cos(\omega t - \delta). \quad . \quad . \quad . \quad (30)$$

It is thus specified, for the present, that the two vibrations have the same frequency. We obtain widely different vibration figures for differing values of the phase difference  $\delta$ . First, let us take  $\delta = 0$ . In this case, the point remains on the straight line

$$\frac{y}{x} = \frac{b}{a},$$

which is the diagonal of a rectangle with sides  $a$  and  $b$ . This motion is called the *linear* form of vibration. If we next take  $\delta$  between 0 and  $\pi/2$ , elimination of  $t$  in (30) gives the equation of the orbit as

$$y = b \frac{x}{a} \cos \delta + b \sqrt{1 - \frac{x^2}{a^2}} \sin \delta, \\ \left( y - \frac{b}{a} x \cos \delta \right)^2 = b^2 \left( 1 - \frac{x^2}{a^2} \right) \sin^2 \delta. \quad . \quad . \quad (31)$$

This is the equation of a second degree curve. From the parametric



form (30), it is evident that neither the  $x$ - nor the  $y$ -co-ordinate can become infinite, i.e. the curve must be an ellipse. Its principal axes may be determined by the methods of analytical geometry. For  $\delta = \pi/2$  we have

$$x = a \cos \omega t, \quad y = b \sin \omega t.$$

This, however, is the familiar parametric representation of the ellipse

$$\frac{x^2}{a^2} + \frac{y^2}{b^2} = 1,$$

whose principal axes are the co-ordinate axes.

The direction in which the ellipse (30) is described may be seen from the parametric representation. If we start at  $t = 0$ , then  $P$  is at its greatest distance from  $O$  in the direction of the  $x$ -axis, i.e. the  $x$  of  $P$  is a maximum. As  $t$  increases,  $x$  decreases, but (if  $\delta$  is less than  $\pi$ )  $y$  increases in the direction of the positive  $y$ -axis. This form of vibration is called a *left-handed elliptic* vibration. If  $\delta$  is allowed to increase to the value  $\pi$ ,

$$x = a \cos \omega t, \quad y = -b \cos \omega t, \quad \frac{y}{x} = -\frac{b}{a}.$$

This is the original linear vibration reflected in the  $y$ -axis. Assigning  $\delta$  the value  $3\pi/2$  gives the same ellipse as for  $\delta = \pi/2$ ; but this curve

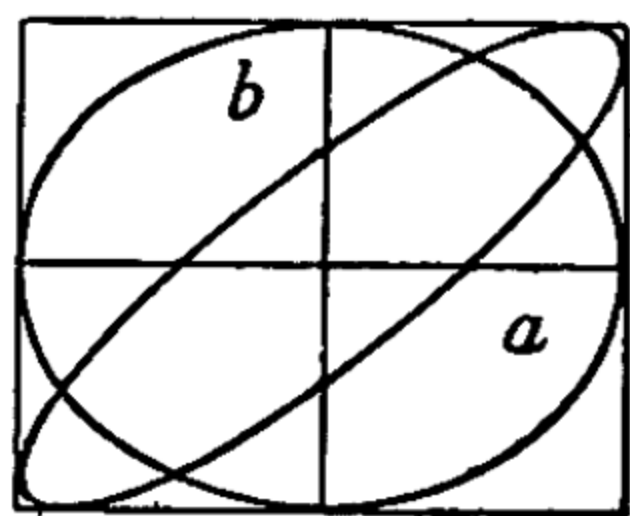


Fig. 3

is traversed in the opposite sense, and we have a case of *right-handed elliptic* motion. We again obtain the linear motion with which we started by putting  $\delta = 2\pi$ . The intermediate ellipses may be drawn with the aid of the above considerations (cf. fig. 3). If the amplitudes  $a$  and  $b$  are equal, we have the special case of motion in a circle. This may be either right-handed or left-handed.

If we employ exponentials instead of trigonometric functions,

$$\left. \begin{aligned} x = z_1 &= ae^{i\omega t}, \\ y = z_2 &= be^{i(\omega t - \delta)}. \end{aligned} \right\} \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad (32)$$

This shows that the ratio of the components of the radius vector is real only when the phase difference is 0 or  $\pi$ , i.e. when the vibration is linear, for

$$\frac{y}{x} = \frac{z_2}{z_1} = \frac{b}{a} e^{-i\delta}. \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad (33)$$

Thus the complex number representation provides a simple criterion of an elliptic vibration: the ratio of the components  $x$  and  $y$  is complex.

If the two frequencies are not equal, the phase difference is con-

stantly changing. For the case where the frequency difference is small compared to the individual frequencies, the phase relation changes slowly, and we can write

$$\left. \begin{aligned} x &= a \cos \omega_1 t, \\ y &= b \cos[\omega_1 t + (\omega_2 - \omega_1)t] = b \cos(\omega_1 t + \epsilon t). \end{aligned} \right\} \quad (34)$$

The phase difference  $\delta$  thus becomes a linear function of the time, and the point  $P$  traverses in smooth succession all the elliptical curves corresponding to the various values of  $\delta$ . The resulting pattern, called a *Lissajous figure*, after its discoverer, has the property of gradually covering the entire rectangle of sides  $a$  and  $b$  (cf. fig. 4). The Lissajous figure is the simplest example of a conditionally periodic system. In this type of system—which is of importance in Celestial Mechanics and, to a certain extent, also in Atomic Mechanics—each co-ordinate of  $P$  is, in fact, a periodic function of the time. The orbit itself is closed only when a condition of commensurability connects the two periods. In this case alone can we speak of a periodic motion.

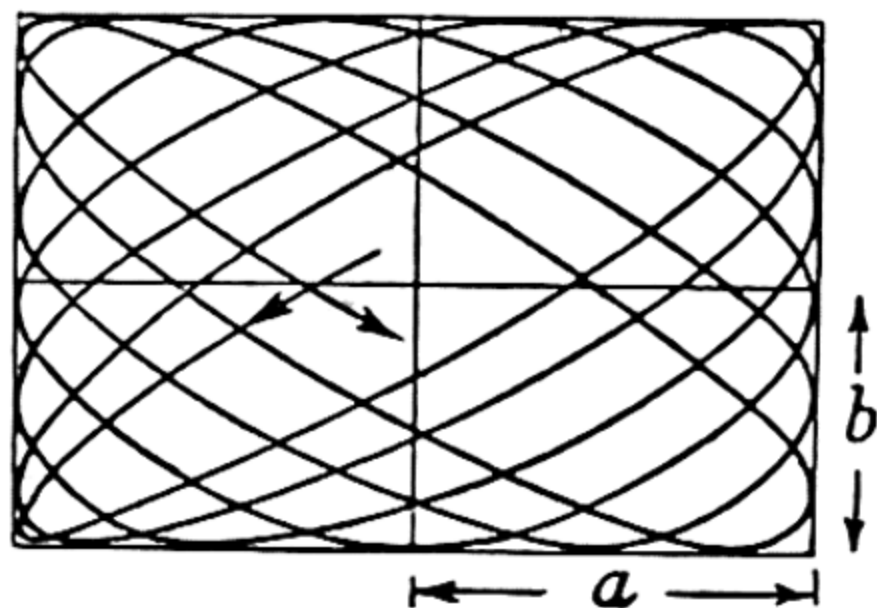


Fig. 4

Since any vector (e.g. the vector representing the electric field strength) may be represented as a radius vector, our considerations are perfectly general, and hold for the motion of the terminus of a vector, as well as for the motion of a material particle. It is further of importance to note that a right-handed and a left-handed circular motion of the same frequency and amplitude always combine to form a linear vibration, whose axis is determined by the phase difference of the two motions. If the left-handed motion is

$$\left. \begin{aligned} x_1 &= a \cos \omega t, \\ y_1 &= a \sin \omega t, \end{aligned} \right\} \quad \cdot \cdot \cdot \cdot \cdot \cdot \quad (35)$$

and the right-handed motion is

$$\left. \begin{aligned} x_2 &= a \cos(\omega t - \delta), \\ y_2 &= -a \sin(\omega t - \delta), \end{aligned} \right\} \quad \cdot \cdot \cdot \cdot \cdot \cdot \quad (36)$$

then a simple trigonometric transformation gives

$$x_1 + x_2 = 2a \cos \frac{\delta}{2} \cos(\omega t - \delta/2). \quad \cdot \cdot \cdot \quad (37)$$

Similarly, 
$$y_1 + y_2 = 2a \sin \frac{\delta}{2} \cos(\omega t - \delta/2). \quad \cdot \cdot \cdot \quad (38)$$

This shows that the  $x$ - and  $y$ -components of the resulting motion do not differ in phase, and the vibration is linear. The direction of the axis is given by division of the two equations:

$$\frac{y_1 + y_2}{x_1 + x_2} = \tan \frac{\delta}{2}. \quad . \quad . \quad . \quad . \quad . \quad (39)$$

This result is still more easily obtained by transferring the circular vibrations to the complex plane, where they are represented by

$$z_1 = ae^{i\omega t} \quad \text{and} \quad z_2 = ae^{-i(\omega t - \delta)}. \quad . \quad . \quad . \quad (40)$$

The sum becomes

$$\begin{aligned} z_1 + z_2 &= ae^{i\delta/2} \{ e^{i(\omega t - \delta/2)} + e^{-i(\omega t - \delta/2)} \} \\ &= 2ae^{i\delta/2} \cos(\omega t - \delta/2). \quad . \quad . \quad . \quad . \quad (41) \end{aligned}$$

This is a vibration taking place in a line making an angle  $\delta/2$  with the real axis.

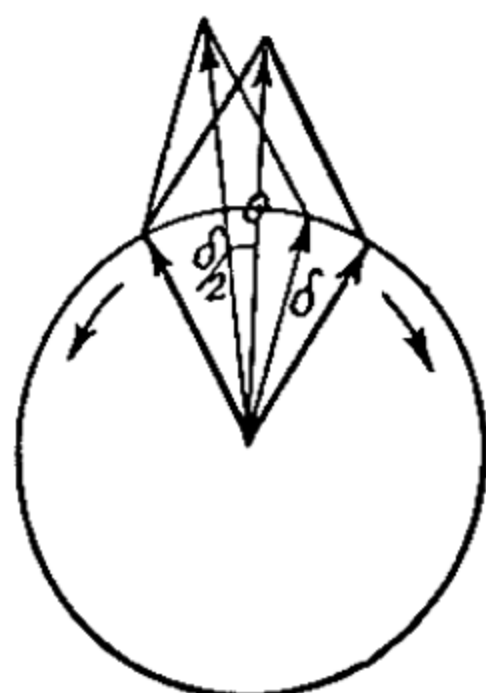


Fig. 5

This method, which is always applicable to circular vibrations, is not to be confused with the vector diagram introduced in § 1 (p. 52). There, only the real part of the complex number had physical significance; in circular vibration, the complex number itself gives the correct picture of the process.

The same result is obtained diagrammatically by combining the instantaneous radius vectors, and again the geometry of the figure gives the angle  $\delta/2$  (fig. 5).

*Ex. 25.* Draw, for several values of the phase difference, the closed vibration figure resulting from combination of an  $x$ - and a  $y$ -vibration having the ratio of frequencies 1 : 2.

## 5. The Propagation of Periodic Disturbances in the form of Waves.

Hitherto we have confined our attention to the particular point in space at which a given vibration is taking place; in what follows we shall consider all points of the region. First we take the simple case in which motion in a certain direction (call this the  $x$ -axis) is independent of the  $y$ - and  $z$ -co-ordinates, i.e. is a function of  $x$  alone. At  $x = 0$  let

$$u(t, 0) = Ae^{i\omega t}. \quad . \quad . \quad . \quad . \quad (42)$$

We call a process a *wave* if the same state of vibration exists at  $x = x$  as at  $x = 0$ , but with a difference of phase corresponding to the finite velocity of propagation  $v$  of the phase.\* That is, at the time  $t$ , the

\*  $v$  is also referred to as the "wave velocity".



same condition obtains at  $x$  as held at 0 at the time  $t - x/v$ . We thus write \*

$$u(t, x) = Ae^{i\omega(t - \frac{x}{v})}. \quad . \quad . \quad . \quad . \quad . \quad (43)$$

In this way  $u$  becomes not only a function of the time, but also of the co-ordinate  $x$ . If we imagine a snapshot of the process to be made at a time  $t_0$ , then the corresponding equation is

$$u(t_0, x) = Ae^{i\omega t_0} e^{-i\omega \frac{x}{v}}. \quad . \quad . \quad . \quad . \quad . \quad (44)$$

The instantaneous value at 0 recurs if  $x$  increases by an integral multiple of

$$\frac{2\pi v}{\omega} = \frac{v}{\nu},$$

for in this case the argument of the imaginary exponential increases by  $2\pi$ , and the function remains unchanged. This distance, which gives the separation of the points of equal phase, is called the *wave-length*  $\lambda$ . From the above considerations,  $\lambda$  obeys the fundamental equation of wave theory:

$$\lambda \nu = v: \quad . \quad . \quad . \quad . \quad . \quad (45)$$

*Wave-length*  $\times$  *Frequency* = *velocity of propagation of the phase*.

A wave of the kind described, in which  $u$  depends, in addition to the time, only upon the co-ordinate of the direction of propagation, is called a *plane wave*, since  $u$  is constant in any plane perpendicular to the direction of propagation. The simple harmonic plane wave is a particular integral of a partial differential equation which is readily deduced. If  $u$  is differentiated twice with respect to  $t$ , and twice with respect to  $x$ , we obtain

$$\frac{\partial^2 u}{\partial t^2} = -\omega^2 Ae^{i\omega(t - x/v)} \quad \text{and} \quad \frac{\partial^2 u}{\partial x^2} = -\frac{\omega^2}{v^2} Ae^{i\omega(t - x/v)} \quad . \quad (46)$$

so that

$$\frac{\partial^2 u}{\partial x^2} = \frac{1}{v^2} \frac{\partial^2 u}{\partial t^2}. \quad . \quad . \quad . \quad . \quad . \quad (47)$$

The general integral of this second order partial differential equation must contain two arbitrary functions. As may be verified readily, any function  $f$  of  $(t - x/v)$  satisfies the equation, but so does any function  $g$  of  $(t + x/v)$ . The latter corresponds to a wave moving in the direction of the negative  $x$ -axis. The general integral is then

$$u = f\left(t - \frac{x}{v}\right) + g\left(t + \frac{x}{v}\right). \quad . \quad . \quad . \quad . \quad (48)$$

\* In the simplest case, as taken here, the amplitude  $A$  is independent of the position in space. Nevertheless,  $A$  may also be a function of  $x$ . Thus, in an absorbing medium the amplitude falls off as  $e^{-\alpha x}$ . Here, however, we consider only *undamped waves* which, in our case of dependence solely on  $x$ , have constant amplitude.



For a plane wave whose wave normal  $\mathbf{n}$  (i.e. the perpendicular to the planes of equal phase) is arbitrarily oriented with respect to the axes, the planes of equal phase have the equation  $\mathbf{rn} = \text{const.}$ , and the plane wave is represented by

$$u = Ae^{i\omega(t - \frac{\mathbf{rn}}{v})} = Ae^{i\omega(t - \frac{x \cos \alpha + y \cos \beta + z \cos \gamma}{v})}. \quad (43')$$

If this equation is differentiated twice with respect to  $t$  and twice with respect to each co-ordinate, we can form the most general differential equation defining an undamped wave, of which (43') is only one particular integral. This equation turns out to be

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 u}{\partial z^2} = \Delta u = \frac{1}{v^2} \frac{\partial^2 u}{\partial t^2}. \quad \dots (49)$$

Another very important special case is that of a spherical wave; here the function depends only on the distance  $r$  from a centre  $O$ . The transformation of  $\Delta u$  to this case—the special case of spherical polar co-ordinates when  $u$  depends on  $r$  only—is not difficult if we use the vector meaning of  $\Delta u$ , viz.  $\text{div grad } u$ . We have

$$\begin{aligned} \text{grad } u(r) &= \frac{du}{dr} \frac{1}{r} \mathbf{r} \\ \text{div grad } u(r) &= \frac{du}{dr} \frac{1}{r} \text{div } \mathbf{r} + \frac{\mathbf{r}\mathbf{r}}{r} \left( \frac{d^2 u}{dr^2} \frac{1}{r} - \frac{du}{dr} \frac{1}{r^2} \right) \\ &= \frac{d^2 u}{dr^2} + \frac{2}{r} \frac{du}{dr}. \end{aligned}$$

Then (49) becomes

$$\frac{\partial^2 u}{\partial r^2} + \frac{2}{r} \frac{\partial u}{\partial r} = \frac{1}{v^2} \frac{\partial^2 u}{\partial t^2}. \quad \dots (50)$$

A particular integral is

$$u = \frac{A}{r} e^{i\omega(t - \frac{r}{v})}, \quad \dots (51)$$

as may be seen readily by substitution, or by writing the equation in the form

$$\frac{\partial^2}{\partial r^2} (ru) = \frac{1}{v^2} \frac{\partial^2}{\partial t^2} (ru).$$

This is a simple harmonic wave spreading equally along radii in all directions from the origin. In contrast with plane waves, the amplitude falls off as  $1/r$ , corresponding to the increase of the wave surfaces. The surfaces of constant phase are given by  $r = \text{const.}$ , and are therefore spheres. The general solution, consisting of one wave emanating from  $O$  and another closing in towards  $O$ , is

$$u = (1/r)\{f(t - r/v) + g(t + r/v)\}. \quad . \quad . \quad . \quad (52)$$

*Ex.* 26. Give the differential equation of a cylindrical wave, and find its integral.

**6. Combination of several Waves having the same direction of Propagation; Linearly and Elliptically Polarized Waves; Group Velocity.**

(a) *Phase velocity independent of the frequency; absence of dispersion*

If  $u$  at  $x = 0$  is a complicated function of the time, representable by a Fourier Series, and if there is no dispersion, then the phenomenon is repeated at every point with a phase difference corresponding to the velocity of propagation. In order to represent such a wave, we have only to replace  $t$  by  $t - x/v$  in the Fourier Series. If the quantity satisfying the wave equation is of the nature of a vector, we speak of pure *longitudinal* waves if this vector is along the direction of propagation, i.e. the  $x$ -axis. In this case everything is cylindrically symmetrical around the  $x$ -axis, and there is no uniquely distinguished direction in the planes perpendicular to this axis. If, on the other hand, the vector has no component in the  $x$ -direction, we speak of pure *transverse* waves. If there is no phase difference at  $O$  between the two components of the vector—i.e. if the vibration is linear, according to § 4 (p. 58)—it remains linear at every point, if the phase velocity is the same for both components. In this case we speak of *linearly polarized waves*, and refer to the plane determined by the directions of propagation and vibration as the plane of vibration. If the vibration is elliptic at  $O$ , it is elliptic, under the same conditions, at all points in space, and we have an *elliptically polarized wave*, of which a special case is the *circularly polarized wave*.

The case in which the phase velocity is somewhat different for the two components is of importance. If we again represent the vectorial point function by means of a radius vector in the planes perpendicular to the  $x$ -axis,

$$y = a \cos \omega \left( t - \frac{x}{v_1} \right)$$

$$z = b \cos \omega \left( t - \frac{x}{v_2} \right) = b \cos \omega \left[ t - \frac{x}{v_1} + x \left( \frac{1}{v_1} - \frac{1}{v_2} \right) \right].$$

Comparison with formulæ (30) (p. 58) shows that the phase difference varies with the position along the  $x$ -axis, independently of the time (cf. fig. 6). Points of equal phase difference, and therefore having the same vibration figure (e.g. a straight line), are apart a distance  $\Lambda$ , given by

$$\omega \Lambda \left( \frac{1}{v_1} - \frac{1}{v_2} \right) = 2\pi \quad \text{or} \quad \frac{1}{\Lambda} = \frac{1}{\lambda_1} - \frac{1}{\lambda_2}. \quad . \quad . \quad (53)$$

Similarly, we obtain a phase difference varying with the co-ordinate

if we have two oppositely directed circularly polarized waves of equal amplitude, but having different velocities of propagation. According

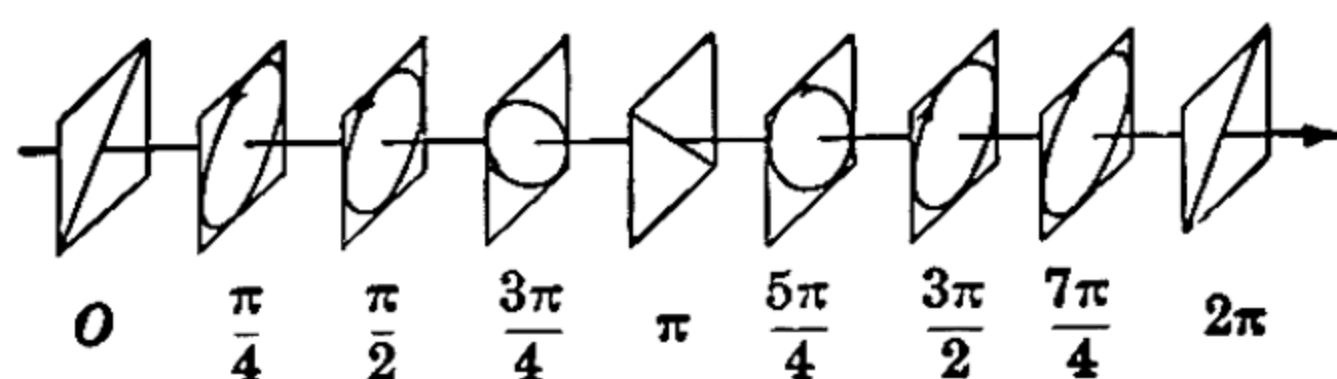


Fig. 6

to § 4 (p. 60), two such vibrations always combine to give a linear vibration, and the position of the axis of vibration is given by the phase difference. If the latter varies from place to place, this signifies a gradual azimuthal rotation of the axis of vibration as we move in the direction of propagation.

(b) *Dispersion; phase velocity dependent upon frequency*

If the partial vibrations composing a complicated vibrational phenomenon are propagated with different velocities, we obtain in general, by superposition of all these vibrations, a non-periodic condition, concerning which little can be said. It is only when the partial vibrations are close together in frequency that the nature of the result is clear. The simplest case is that of two slightly different vibrations of equal amplitude which produce a modulated vibration by means of beats. As long as there is no dispersion, the modulation (in radio, the audio-frequency note) is, of course, transmitted with the same velocity as the carrier wave. If, however, dispersion exists, we obtain another velocity—the *group velocity*—as the following simple computation shows:

Let  $\omega_1 = \omega + \Delta\omega$  and  $v_1 = v + \frac{dv}{d\omega} \Delta\omega$ .

Then the modulated wave becomes

$$\begin{aligned} F(x, t) &= \cos \omega_1 \left( t - \frac{x}{v_1} \right) + \cos \omega \left( t - \frac{x}{v} \right) \\ &= 2 \cos \left\{ \frac{\omega_1}{2} \left( t - \frac{x}{v_1} \right) + \frac{\omega}{2} \left( t - \frac{x}{v} \right) \right\} \\ &\quad \cos \left\{ \frac{\omega_1}{2} \left( t - \frac{x}{v_1} \right) - \frac{\omega}{2} \left( t - \frac{x}{v} \right) \right\}. \end{aligned}$$

We may set  $\omega_1 = \omega$  and  $v_1 = v$  in the sum, without appreciable error. Then

$$F(x, t) = 2 \cos \omega \left( t - \frac{x}{v} \right) \cos \left\{ \frac{\Delta\omega}{2} t - \frac{1}{2} \left( \frac{\omega_1}{v_1} - \frac{\omega}{v} \right) x \right\}.$$



Neglecting higher terms,

$$\frac{\omega_1}{v_1} - \frac{\omega}{v} = \Delta \frac{\omega}{v} = \left( \frac{1}{v} - \frac{\omega}{v^2} \frac{dv}{d\omega} \right) \Delta \omega.$$

Thus one obtains

$$F(x, t) = 2 \cos \frac{\Delta \omega}{2} \left( t - \frac{x}{V} \right) \cos \omega \left( t - \frac{x}{v} \right),$$

where, for brevity, we have set

$$\frac{1}{v} - \frac{\omega}{v^2} \frac{dv}{d\omega} = \frac{d \frac{\omega}{v}}{d\omega} = \frac{d \frac{v}{\omega}}{dv} = \frac{1}{V}.$$

The meaning of this result is as follows: We have a wave of angular frequency  $\omega$ ; its amplitude is subject to a wave-like periodic variation \* of angular frequency  $\Delta \omega$ . While the phase of the wave is propagated with velocity  $v$ , the velocity of propagation of the modulation has the value  $V$ , determined by

$$\frac{1}{V} = \frac{d \frac{v}{\omega}}{dv} \quad \cdot \cdot \cdot \cdot \cdot \cdot \cdot \quad (54)$$

It is useful to express the group velocity also in terms of  $\lambda$ . Since  $v/\omega = \lambda$ , (54) may be written

$$V = \frac{dv}{d(1/\lambda)} = -\lambda^2 \frac{dv}{d\lambda},$$

and since  $v\lambda = v$ ,

$$\lambda \frac{dv}{d\lambda} + v = \frac{dv}{d\lambda}.$$

Substituting the value of  $\lambda(dv/d\lambda)$  into the above, we obtain

$$V = v - \lambda \frac{dv}{d\lambda} \quad \cdot \cdot \cdot \cdot \cdot \cdot \cdot \quad (54')$$

The velocity  $V$  is, in general, identical with the signal velocity, i.e. the velocity of propagation of a signal. If we have a vibration at the origin  $O$  starting at time  $t = 0$  and terminating at  $t = T$ , the function which represents it is in general non-periodic, but may be represented by a continuous sequence of harmonic functions, according to § 2 (p. 56). If the duration  $T$  of the original vibration is small, we obtain a narrow band of side frequencies, for which, if the dispersion is not too great, the derivative  $dv/d\omega$  may be regarded as constant. In the same way as a modulated wave results from superposition of two slightly different primary waves, a vibration

\* Because the algebraic sign of the amplitude is of no concern, the frequency is not  $\Delta \omega/2$  but  $\Delta \omega$  (cf. p. 56).

lasting for a limited time at any one point results in this case by superposition of all the partial frequencies. The velocity of propagation of the limited wave train is then given by  $V$ . Since signals can be given only by initiating or terminating a vibration process, and not by individual crests or troughs of an unlimited wave,  $V$  is the velocity of propagation of acoustical or optical signals, as obtained by measurement.

In the region of anomalous dispersion, the signal velocity, i.e. the velocity of propagation of the energy, no longer coincides with the group velocity  $V$ , which loses its significance in this instance.

*Ex. 27.* How must the phase velocity  $v$  depend on  $\nu$ , in order that the group velocity  $V$  be proportional to the reciprocal of the phase velocity?

## 7. Combination of Waves having the same Frequency but different directions of Propagation. Standing Waves.

### (a) Superposition of two waves moving in opposite directions

Let

$$u_1 = Ae^{i\omega(t-x/v)}$$

represent a wave travelling in the direction of the positive  $x$ -axis, and let

$$u_2 = Ae^{i\omega(t+x/v)+i\delta}.$$

represent a wave of the same amplitude and frequency travelling in the direction of the negative  $x$ -axis. The sum is

$$\begin{aligned} u_1 + u_2 &= Ae^{i(\omega t + \frac{\delta}{2})} \left\{ e^{-i(\frac{\omega x}{v} + \frac{\delta}{2})} + e^{i(\frac{\omega x}{v} + \frac{\delta}{2})} \right\} \\ &= 2A \cos \left( \frac{\omega x}{v} + \frac{\delta}{2} \right) e^{i(\omega t + \delta/2)}. \quad \dots \quad (55) \end{aligned}$$

This result is to be interpreted as follows:

$$Ae^{i(\omega t + \delta/2)}$$

represents a vibration having the *same phase* all through space. The factor

$$\cos \left( \frac{\omega x}{v} + \frac{\delta}{2} \right)$$

indicates that the amplitude is a periodic function of  $x$ . The period is  $v/\nu = \lambda$ . The chief characteristic of a wave, viz. the finite velocity of propagation of the phase, is entirely absent. For this reason, the phenomenon is called a *standing wave*. If the two amplitudes are unequal, there exists, in addition to the standing wave, a wave travelling in the direction of the stronger partial wave, whose amplitude is the difference of the amplitudes of the components.

(b) *Superposition of two plane waves meeting at an angle*

A case arising especially frequently in optics is that of a plane wave reflected from a plane surface which we shall take to be the  $xy$ -plane of a co-ordinate system. Call the angle between the wave-normal and the  $z$ -axis  $\alpha$ . If we let the wave-normal  $\mathbf{n}$  fall in the  $yz$ -plane, as in fig. 7, then, by the laws of optics and acoustics, the normal to the reflected wave,  $\mathbf{n}'$ , will also lie in the  $yz$ -plane, and will also make an angle  $\alpha$  with the  $z$ -axis. For the incident wave, the surfaces of equal phase are the planes perpendicular to  $\mathbf{n}$ . These planes have the equation  $\mathbf{r}\mathbf{n} = \text{const.}$  The equation of a wave moving in the direction of the unit vector

$$\mathbf{n} = (\cos \alpha)\mathbf{i} + (\cos \beta)\mathbf{j} + (\cos \gamma)\mathbf{k}$$

is thus  $u = Ae^{i\omega(t - \frac{\mathbf{r}\mathbf{n}}{v})} = Ae^{i\omega(t - \frac{x \cos \alpha + y \cos \beta + z \cos \gamma}{v})}$ .

From this equation there follows at once, by putting  $\mathbf{n} = \mathbf{i}$ , the pre-

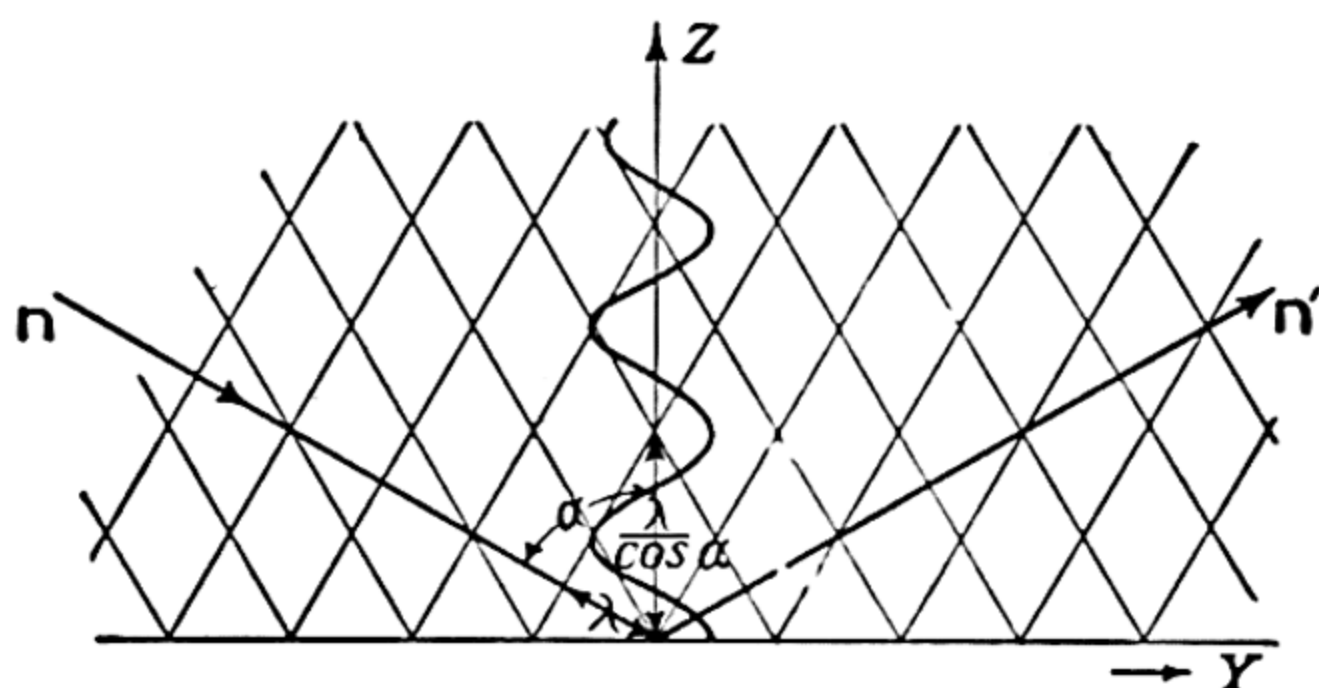


Fig. 7

vious form of a wave moving in the  $x$ -direction. In our case we have (cf. fig. 7),

$$\mathbf{n} = (\sin \alpha)\mathbf{j} - (\cos \alpha)\mathbf{k}; \quad \mathbf{n}' = (\sin \alpha)\mathbf{j} + (\cos \alpha)\mathbf{k}.$$

By superposition of the two waves we obtain

$$\begin{aligned} u_1 + u_2 &= Ae^{i\omega(t - \frac{y \sin \alpha - z \cos \alpha}{v})} + Ae^{i\omega(t - \frac{y \sin \alpha + z \cos \alpha}{v})} \\ &= 2Ae^{i\omega(t - \frac{y \sin \alpha}{v})} \cos\left(\frac{\omega z \cos \alpha}{v}\right). \end{aligned}$$

This is a wave travelling in the direction of the  $y$ -axis, the wave-front being modulated with the period  $\lambda/\cos \alpha$ . The phase velocity of this combination wave is  $v/\sin \alpha$ . For  $\alpha = 0$  we again have the case of a standing wave.



## CHAPTER III

### SELECTED TOPICS IN THE THEORY OF FUNCTIONS OF A COMPLEX VARIABLE

#### 1. Conformal Mapping of one Plane on another.

Since many problems of theoretical physics may be solved particularly simply with the aid of the Theory of Complex Functions, we give here a rapid account of some of the most useful results of function theory. For rigorous deductions of these results, however, the reader should consult textbooks on the subject.

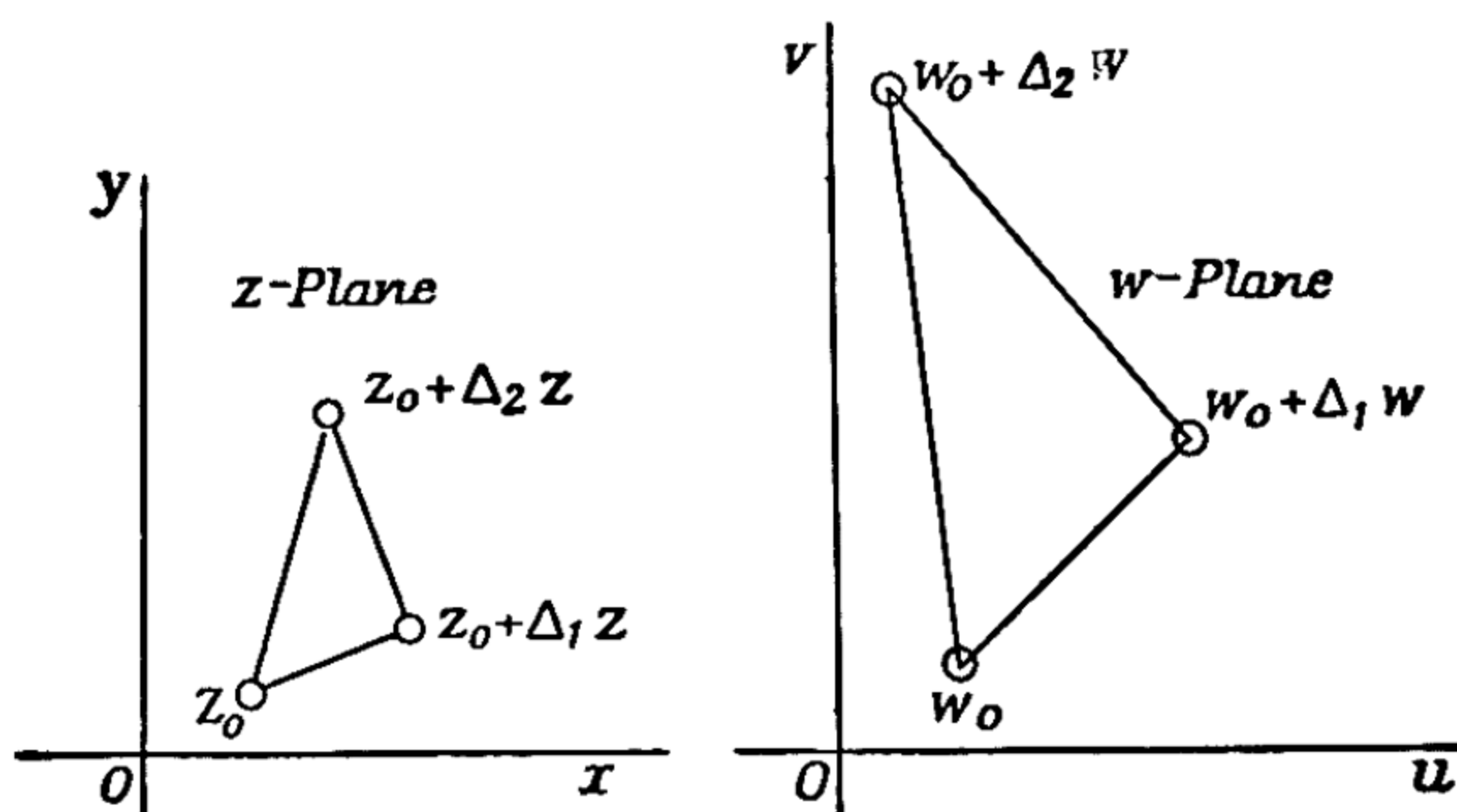


Fig. 1

If the argument of a function, e.g. the sine function, is a complex number  $z = x + iy$ , then, in general, the function will also be a complex number  $w = u + iv$ . If we represent the complex number  $z$  by a point in a plane—the Gauss plane or Argand diagram—and the complex number  $w$  by a point in a second plane, then the function  $f(z)$  establishes a correspondence of the points of the  $w$ -plane with those of the  $z$ -plane, and we say that one plane is *mapped* on the other. The function  $f(z)$  is said to be *monogenic* at those points where  $w$  is finite and where the limit

$$f'(z) = \lim_{\Delta z \rightarrow 0} \frac{f(z + \Delta z) - f(z)}{\Delta z} \quad \dots \quad (1)$$

is finite and independent of the way in which the point  $z$  is approached

as  $\Delta z$  tends to zero. We now state a theorem: Within a region in which  $w$  is regular, the  $z$ -plane may be mapped on the  $w$ -plane, with the exception of those points where  $f'(z) = 0$ , by means of a function  $w = f(z)$ , in such a way that corresponding small portions of each plane are similar and angles are preserved (conformal mapping). For a proof, we consider a small triangle in the  $z$ -plane, determined by the points  $z_0$ ,  $z_0 + \Delta_1 z$  and  $z_0 + \Delta_2 z$  (fig. 1). These points correspond to the points  $w_0$ ,  $w_0 + \Delta_1 w = w_0 + f'(z_0) \Delta_1 z$  and  $w_0 + \Delta_2 w = w_0 + f'(z_0) \Delta_2 z$ , in the  $w$ -plane. Thus, the quantities  $\Delta_1 w$  and  $\Delta_2 w$  are obtained by multiplying the complex numbers  $\Delta_1 z$  and  $\Delta_2 z$  by the same complex number  $f'(z_0)$ . But this means that  $\Delta_1 z$  and  $\Delta_2 z$  are turned through the same angle and elongated in the same ratio.\* The similarity of the two triangles is thus demonstrated; all angles

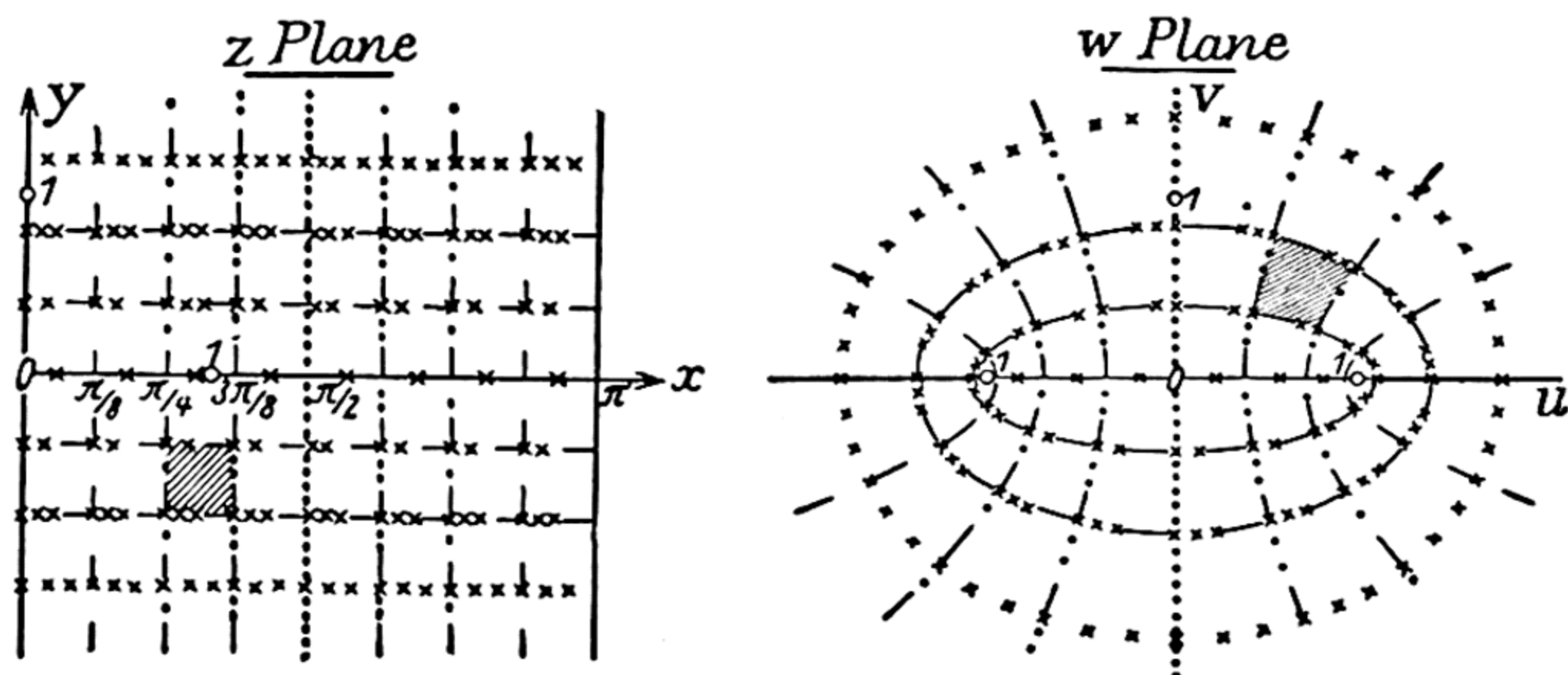


Fig. 2

are preserved in the mapping. The type of mapping represented by the function  $w = f(z)$  is therefore termed *conformal*. The lines  $x = \text{const.}$  and  $y = \text{const.}$  in the  $z$ -plane form a rectilinear rectangular co-ordinate system. On account of the preservation of angles, these straight lines correspond to two families of mutually perpendicular curves in the  $w$ -plane. As an example, consider the function

$$\begin{aligned} u + iv = w = \cos z &= \frac{1}{2}(e^{iz} + e^{-iz}) = \frac{1}{2}e^{ix-y} + \frac{1}{2}e^{-ix+y} \\ &= \cosh y \cos x - i \sinh y \sin x, \end{aligned}$$

that is

$$u = \cosh y \cos x, \quad v = -\sinh y \sin x.$$

Elimination of  $x$  yields

$$\frac{u^2}{\cosh^2 y} + \frac{v^2}{\sinh^2 y} = 1.$$

\* We see at once that the multiplication of two complex numbers represents a rotation and a change of length. If  $z_1 = \rho_1 e^{i\phi_1}$  and  $z_2 = \rho_2 e^{i\phi_2}$ , then  $z_1 z_2 = \rho_1 \rho_2 e^{i(\phi_1 + \phi_2)}$ .

Similarly by eliminating  $y$ ,

$$\frac{u^2}{\cos^2 x} - \frac{v^2}{\sin^2 x} = 1.$$

The curves  $x = \text{const.}$  thus become, in the  $w$ -plane, hyperbolas with foci at  $-1$  and  $+1$ ; the curves  $y = \text{const.}$  become ellipses having the same foci (fig. 2). It is known from analytic geometry that these confocal conics form an orthogonal system. The strip of the  $z$ -plane between  $x = 0$  and  $x = \pi$  maps itself into the entire  $w$ -plane. To represent the entire  $z$ -plane requires a  $w$ -plane of infinitely many sheets.

*Ex. 28.* Discuss the transformation  $w = z + 1/z$ .

## 2. The Cauchy-Riemann Conditions and the Differential Equation of Laplace.

The real part  $u$  and the imaginary part  $v$  of  $w = u + iv$  are functions of the two variables  $x$  and  $y$ . Nevertheless, two arbitrary functions  $u(x, y)$  and  $v(x, y)$  cannot, in general, be considered to be the real and imaginary parts of a function of a complex variable. If the functions  $u$  and  $v$  originate in a complex function, they satisfy certain special conditions. If we denote by  $f'(z)$  the derivative of the function  $f$  with respect to its argument, then, since  $u + iv = f(x + iy) = f(z)$ ,

$$\frac{\partial u}{\partial x} + i \frac{\partial v}{\partial x} = f'(z) \frac{\partial z}{\partial x} = f'(z);$$

$$\frac{\partial u}{\partial y} + i \frac{\partial v}{\partial y} = f'(z) \frac{\partial z}{\partial y} = i f'(z).$$

Comparison gives the *Cauchy-Riemann Differential Equations*

$$\left. \begin{aligned} \frac{\partial u}{\partial x} &= \frac{\partial v}{\partial y}, \\ \frac{\partial u}{\partial y} &= -\frac{\partial v}{\partial x}. \end{aligned} \right\} \dots \dots \dots (2)$$

These relationships must hold between  $u$  and  $v$ , but neither of the two functions may be chosen at will. If we differentiate the first of the above equations with respect to  $x$ , the second with respect to  $y$ , and add, we obtain *Laplace's Differential Equation*

$$\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 u}{\partial y^2} = \Delta u = 0. \quad \dots \dots \dots (3)$$

Similarly, by differentiating the first with respect to  $y$  and the second with respect to  $x$ , and subtracting.



$$\frac{\partial^2 v}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} = \Delta v = 0. \quad . . . . . (3')$$

Hence both the real and the imaginary parts of the complex function  $w$  satisfy Laplace's equation:  $\Delta u = 0$ ,  $\Delta v = 0$ .

### 3. Line Integrals in the Gauss Plane; the Cauchy Integral Theorem.

Form the integral  $\int_{P_0}^{P_1} f(z) dz$  between two points  $P_0$  and  $P_1$  of the curve  $C$  in the  $z$ -plane (fig. 3). Cauchy's theorem is: The integral along a second curve  $C'$ , between the same points, has the same value if the function  $f(z)$  is monogenic at all points in the region bounded by the two curves; or, what is the same thing, *the contour integral  $\oint f(z) dz$  vanishes if the function is monogenic in the enclosed region.* To prove this statement, we separate real and imaginary parts:

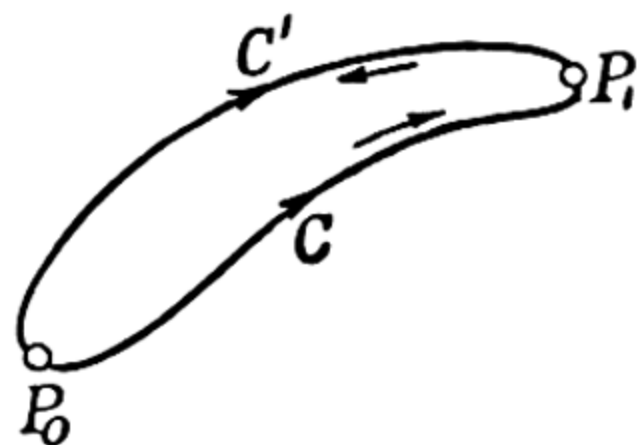


Fig. 3

$$\begin{aligned} \oint (u + iv)(dx + idy) &= \oint (u dx - v dy) + i \oint (v dx + u dy) \\ &= \oint \mathbf{F} d\mathbf{s} + i \oint \mathbf{F}' d\mathbf{s}. \end{aligned}$$

The first integral represents a plane line integral  $\oint \mathbf{F} d\mathbf{s}$  if we put  $F_x = u$  and  $F_y = -v$ . But by Stokes's Theorem (p. 27), since  $d\mathbf{S}$  has a  $z$ -component only,

$$\oint \mathbf{F} d\mathbf{s} = \int \text{curl } \mathbf{F} d\mathbf{S} = \iint \left( -\frac{\partial v}{\partial x} - \frac{\partial u}{\partial y} \right) dx dy,$$

and similarly, for the second integral,

$$\oint \mathbf{F}' d\mathbf{s} = \int \text{curl } \mathbf{F}' d\mathbf{S} = \iint \left( \frac{\partial u}{\partial x} - \frac{\partial v}{\partial y} \right) dx dy.$$

However, by the Cauchy-Riemann equations (2), the integrands in both double integrals vanish identically, which proves the theorem. Further, the following theorem may be proved: If the line integral along a closed curve  $C$  does not vanish, but has a finite value, on account of the existence of a singularity of the function in the region enclosed by  $C$ , then the same value is obtained by integrating along another closed curve  $C'$ , provided the function is monogenic in the region between the two curves. The two curves  $C$  and  $C'$  (fig. 4) form the boundary of a region to which Stokes's Theorem may be applied, since the function  $f$  (and therefore also  $u$  and  $v$ ) are regular within

this region.\* It is to be remembered, however, that in applying Stokes's Theorem, the boundary curve is to be traversed in such direction that the bounded surface is always on the left. Hence the direction in which the inner surface is described must be reversed; or, if the curves are traversed in the same sense,

$$\oint_C - \oint_{C'} = 0.$$

If there is but *one* singular point  $z_0$ , we may therefore integrate around a small circle of radius  $\rho$ , with centre at  $z_0$ . In the greater number of cases, the singularity at  $z_0$  consists in the function becoming

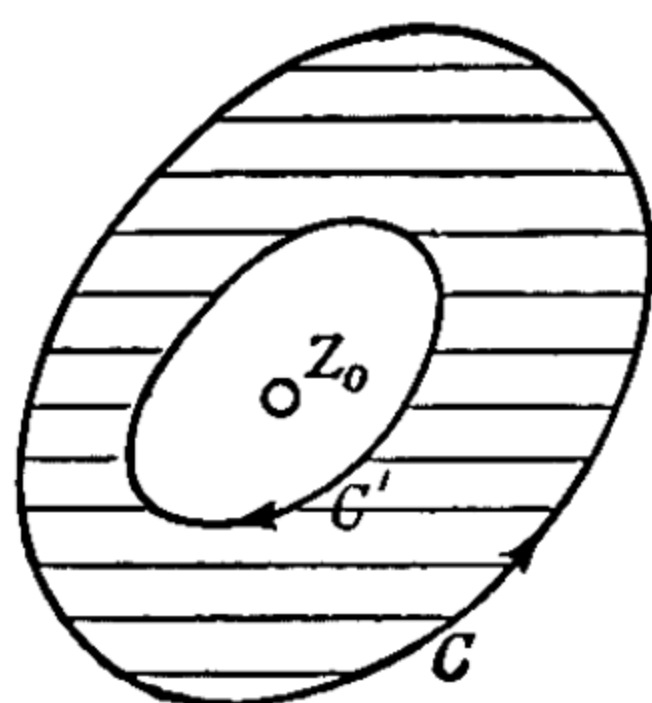


Fig. 4

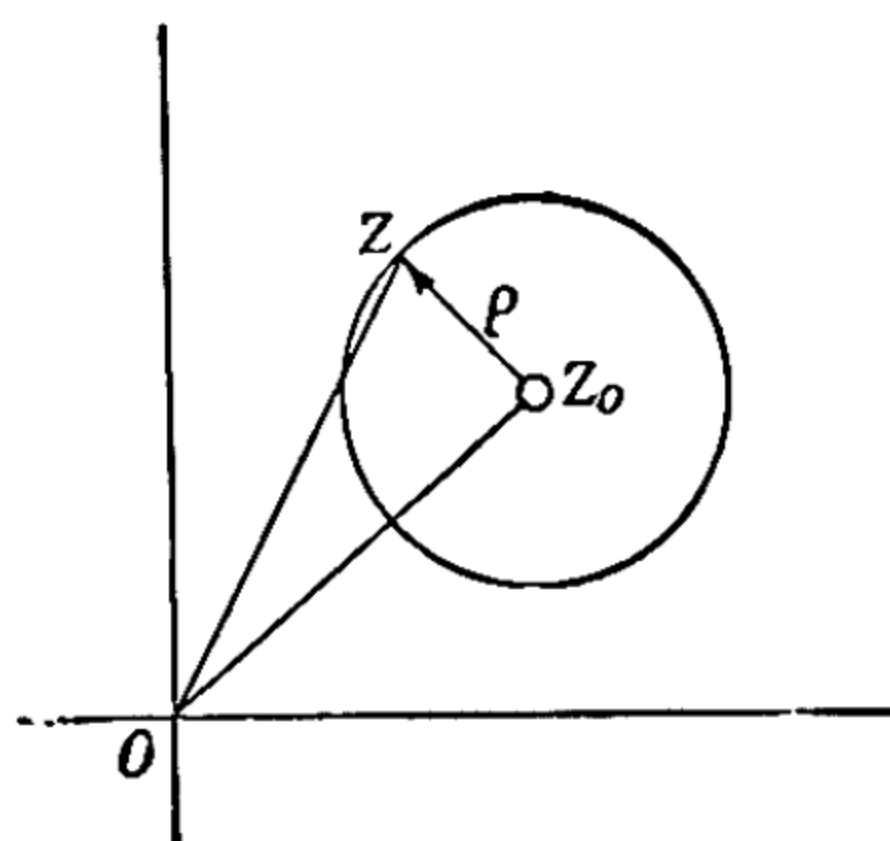


Fig. 5

infinite to a certain definite order  $n$ ; i.e. a power series development at  $z_0$  contains terms  $(z - z_0)^{-1}, \dots (z - z_0)^{-n}$ . It may be shown at once that the value of the integral depends only upon the term  $(z - z_0)^{-1}$ , for if we put (cf. fig. 5)

$$z - z_0 = \rho e^{i\phi}, \quad dz = i\rho e^{i\phi} d\phi,$$

then

$$\oint (z - z_0)^n dz = \int_0^{2\pi} i\rho^{n+1} e^{(n+1)i\phi} d\phi = \frac{\rho^{n+1}}{n+1} \left[ e^{i(n+1)\phi} \right]_0^{2\pi}.$$

On account of the periodicity of the complex exponential function, the integrals vanish for  $n \neq -1$ , but for  $n = -1$ ,

$$\oint \frac{dz}{z - z_0} = \int_0^{2\pi} \frac{i\rho e^{i\phi} d\phi}{\rho e^{i\phi}} = 2\pi i. \quad \dots \dots (4)$$

If the path of integration encloses a singularity  $z_0$  of the function  $f$ , the

\* From a strict, mathematical viewpoint it is allowable to apply the Stokes's Theorem only after joining the two curves by a "cut". In forming the line integral this cut will be traversed twice in opposite directions, hence contributes nothing to the integral. By inserting the cut we get a single, continuous, bounding curve that can be traversed in one operation.

value of the complete line integral is  $2\pi i$  times the coefficient of the term  $(z - z_0)^{-1}$  of the power series development (in powers of  $z - z_0$ ), which holds in the neighbourhood of  $z_0$ . This coefficient is called the *residue* of the function at  $z_0$ . If several singularities are enclosed, it is readily shown that the value of the integral is equal to the sum of the separate residues multiplied by  $2\pi i$ . If the point  $z = \infty$  is enclosed, the substitution  $z = 1/t$ ,  $dz = -dt/t^2$  is made, which reduces the point at infinity to the zero-point. Again the essential coefficient is that of the term in  $z^{-1}$ , for this yields the term in  $1/t$  upon making the above substitution.

*Ex. 29.* Using the Cauchy Theorem, evaluate the integral

$$I = \int_0^{2\pi} \frac{d\phi}{1 + \epsilon \cos \phi}$$

by expressing  $\cos \phi$  in terms of  $e^{i\phi}$  and introducing  $e^{i\phi} = z$  as a new variable.



## CHAPTER IV

### THE FUNDAMENTAL PROBLEM OF THE CALCULUS OF VARIATIONS AND ITS SOLUTION

#### 1. Statement of the Problem of the Calculus of Variations.

The question "For what values of the independent variable has a *given* function  $y = f(x)$  a turning value?" is treated in elementary differential calculus. A necessary condition is the vanishing of the first derivative. Similarly, a necessary condition for an extreme value—i.e. a maximum or minimum, or turning value, or extremum—of a function of several variables is the vanishing of all partial derivatives of the first order  $\partial f/\partial x$ ,  $\partial f/\partial y$ . . . . The following question is much more difficult to answer: "Given a definite integral whose integrand is a function of  $x$ ,  $y$  and of the first derivative  $y'$ :

$$I = \int_{x_0}^{x_1} F(x, y, y') dx;$$

for what function  $y(x)$  is the value of the integral a maximum or a minimum?" In contrast with the simple extreme-value problem of differential calculus, the function  $y(x)$  is not known here, but is to be determined in such way that the integral is an extremum. In geometry and in physics problems of this type meet us at every turn. The simplest example is furnished by the question: "What is the shortest curve which can be drawn joining two given points?" In a plane, the answer may be given immediately: "The straight line." But if the two points and their connecting curve are to lie on a given arbitrary surface, then the analytic equation of this curve—which is called a *geodesic*—is found only by solution of the above extreme-value problem, which is the fundamental problem of the Calculus of Variations.

#### 2. Derivation of the Euler-Lagrange Differential Equation.

We shall now show how the extremal problem of the calculus of variations may be reduced to the determination of the extreme value of a known function. For this purpose we consider functions  $\tilde{y}$  of  $x$ , which are "neighbouring" functions to  $y$ . The function  $\tilde{y}$  is arrived at as follows: Let  $\epsilon$  be a small quantity, and let  $\eta(x)$  be an arbitrary

function of  $x$ , continuous, along with its first two derivatives, in the range of integration. Then we introduce into the integral, in place of  $y$  and  $y'$ , the neighbouring functions  $\tilde{y} = y + \epsilon\eta$  and  $\tilde{y}' = y' + \epsilon\eta'$ . We stipulate, however, that these comparison functions  $\tilde{y}$  coincide with the function  $y(x)$  at the end points of the range of integration (fig. 1). We must therefore require of the arbitrary function  $\eta$  that it vanish at the ends of the interval; for if the various curves connecting two given points are compared, an obvious condition is that all curves to be compared shall pass through these two points, and that no curves which do not pass through these points shall be considered. If we substitute the neighbouring functions in the integral  $I$ , the integral becomes a function of  $\epsilon$ . We then require that  $y(x)$  make the integral an extremum, i.e. the function  $I(\epsilon)$  must have an extreme value for  $\epsilon = 0$ .

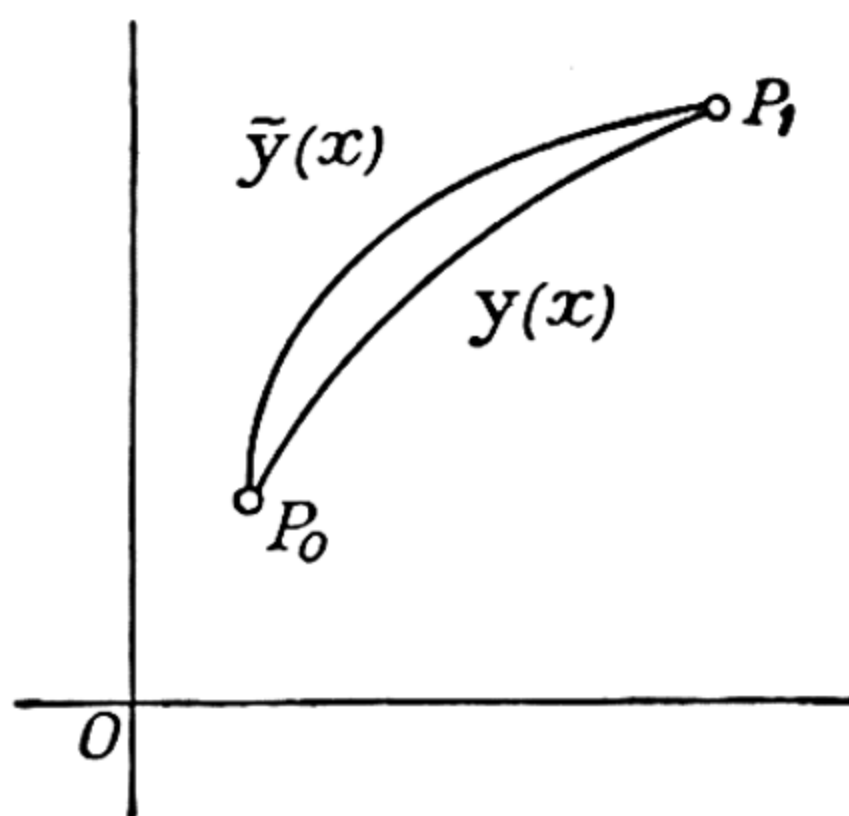


Fig. 1

$$I(\epsilon) = \int_{x_0}^{x_1} F(x, y + \epsilon\eta, y' + \epsilon\eta') dx = \text{extremum for } \epsilon = 0. \quad (1)$$

This gives us a simple method of determining the extreme value of a given integral. The condition is

$$\left( \frac{dI}{d\epsilon} \right)_{\epsilon=0} = 0. \quad . . . . . (2)$$

We expand the integrand function  $F$  in a *Taylor's Series*, according to powers of  $\epsilon$ , and obtain

$$I(\epsilon) = \int_{x_0}^{x_1} \left\{ F(x, y, y') + \epsilon\eta \frac{\partial F}{\partial y} + \epsilon\eta' \frac{\partial F}{\partial y'} + \text{terms in } \epsilon^2, \epsilon^3 \dots \right\} dx. \quad (3)$$

If we differentiate with respect to  $\epsilon$  under the integral sign, as is allowable, we obtain

$$\frac{dI}{d\epsilon} = \int_{x_0}^{x_1} \left\{ \eta \frac{\partial F}{\partial y} + \eta' \frac{\partial F}{\partial y'} + \text{terms in } \epsilon, \epsilon^2 \dots \right\} dx. \quad . (4)$$

This expression is to vanish for  $\epsilon = 0$ . The terms in  $\epsilon$  do this of themselves. Thus there remains the condition for an extremum:

$$\int_{x_0}^{x_1} \left( \eta \frac{\partial F}{\partial y} + \eta' \frac{\partial F}{\partial y'} \right) dx = 0. \quad . . . . . (5)$$

The second term may be transformed by integration by parts:

$$\int_{x_0}^{x_1} \eta' \frac{\partial F}{\partial y'} dx = \left[ \eta \frac{\partial F}{\partial y'} \right]_{x_0}^{x_1} - \int_{x_0}^{x_1} \eta \frac{d}{dx} \frac{\partial F}{\partial y'} dx.$$

The first term vanishes since  $\eta$  is zero at the limits. Then

$$0 = \left( \frac{dI}{d\epsilon} \right)_{\epsilon=0} = \int_{x_0}^{x_1} \eta \left\{ \frac{\partial F(x, y, y')}{\partial y} - \frac{d}{dx} \frac{\partial F(x, y, y')}{\partial y'} \right\} dx. \quad (6)$$

On account of the fact that  $\eta$  is perfectly arbitrary, the integral can vanish only if the bracketed expression in the integrand is zero. This gives the *Euler-Lagrange Differential Equation*:

$$\frac{d}{dx} \frac{\partial F(x, y, y')}{\partial y'} - \frac{\partial F(x, y, y')}{\partial y} = 0, \quad . . . \quad (7)$$

which must be satisfied by  $y$ , if  $y$  is to make the integral an extremum. The essentially difficult investigation of the nature of the extremum (whether maximum or minimum) seldom arises in physics, and need not be treated here. On the other hand, the case where  $F$  is a function of several dependent variables  $y_k$  and their derivatives is of great importance. In analogy with the above, we introduce as neighbouring functions

$$\tilde{y}_1 = y_1 + \epsilon_1 \eta_1; \dots \tilde{y}_k = y_k + \epsilon_k \eta_k, \dots \tilde{y}_f = y_f + \epsilon_f \eta_f,$$

where the  $\eta_k$  are again to vanish at the limits of the integral. The integral then becomes a function of the  $f$  variables  $\epsilon_1 \dots \epsilon_f$ . The condition for an extremum is, of course, the vanishing of the partial derivatives

$$\frac{\partial I(\epsilon_1, \epsilon_2, \dots, \epsilon_f)}{\partial \epsilon_k} = \int_{x_0}^{x_1} \eta_k \left( \frac{\partial F}{\partial y_k} - \frac{d}{dx} \frac{\partial F}{\partial y'_k} \right) dx$$

when

$$\epsilon_1 = \epsilon_2 = \dots = \epsilon_f = 0.$$

It follows as before that the coefficient of each of the functions  $\eta$  within the integral sign must vanish, i.e.

$$\frac{d}{dx} \frac{\partial F}{\partial y'_k} - \frac{\partial F}{\partial y_k} = 0 \quad (k = 1, 2, \dots, f).$$

The Euler-Lagrange equation therefore holds for each of the dependent variables.

The difference

$$I(\epsilon) - I(0) = \epsilon \frac{dI}{d\epsilon}$$



for  $\epsilon$  small, is usually termed the *variation* of the integral, and is written

$$\delta \int_{x_0}^{x_1} F(x, y, y') dx.$$

We then write, instead of

$$\int_{x_0}^{x_1} F(x, y, y') dx = \text{extremum},$$

the equation expressing the vanishing of the variation:

$$\delta \int_{x_0}^{x_1} F(x, y, y') dx = \int_{x_0}^{x_1} \delta F(x, y, y') dx = 0,$$

for which the Euler-Lagrange equation must be satisfied, just as for (1) above.

We call  $\delta F$  the *variation* of  $F$ . An important conclusion may be drawn from this formulation of the variation problem. Let  $F$  be the total derivative of a function  $G(x, y)$ :

$$F(x, y, y') = \frac{d}{dx} G(x, y).$$

In this case the integral may be evaluated immediately:

$$\int_{x_0}^{x_1} F dx = \int_{x_0}^{x_1} \frac{dG}{dx} dx = G(x_1, y_1) - G(x_0, y_0).$$

Since the varied function coincides with the original function at the limits, we have for *every* connexion between  $y$  and  $x$ :

$$\delta \int_{x_0}^{x_1} F dx = \delta G(x_1, y_1) - \delta G(x_0, y_0) = 0.$$

That is, any arbitrary functional relationship  $y(x)$  satisfies the condition that the variation of the integral shall vanish, and the extremal problem no longer has any meaning.

*Ex. 30.* Prove the following: If, in the integral whose extreme value is sought, the integrand  $F$  does not contain  $x$  explicitly, then a first integral is

$$F - y' \frac{\partial F}{\partial y'} = \text{const.}$$

**PART II**  
**MECHANICS**





## CHAPTER V

### THE MECHANICS OF A SINGLE PARTICLE

#### 1. The Fundamental Concepts of Kinematics.

We shall concern ourselves, in this first chapter on mechanics, with the motion of bodies whose extension in space may be neglected, and which may therefore be represented schematically by means of particles, or points endowed with mass (cf. § 2, p. 84). The position of a particle is indicated by a radius vector  $\mathbf{r}$  drawn from a fixed point  $O$ . If necessary, we may resolve this vector into components. At the time  $t$ , let the particle  $P$  be at a point in space specified by the vector  $\mathbf{r}$ . At the time  $t + dt$  the vector is  $\mathbf{r} + d\mathbf{r}$  (fig. 1). The *velocity*

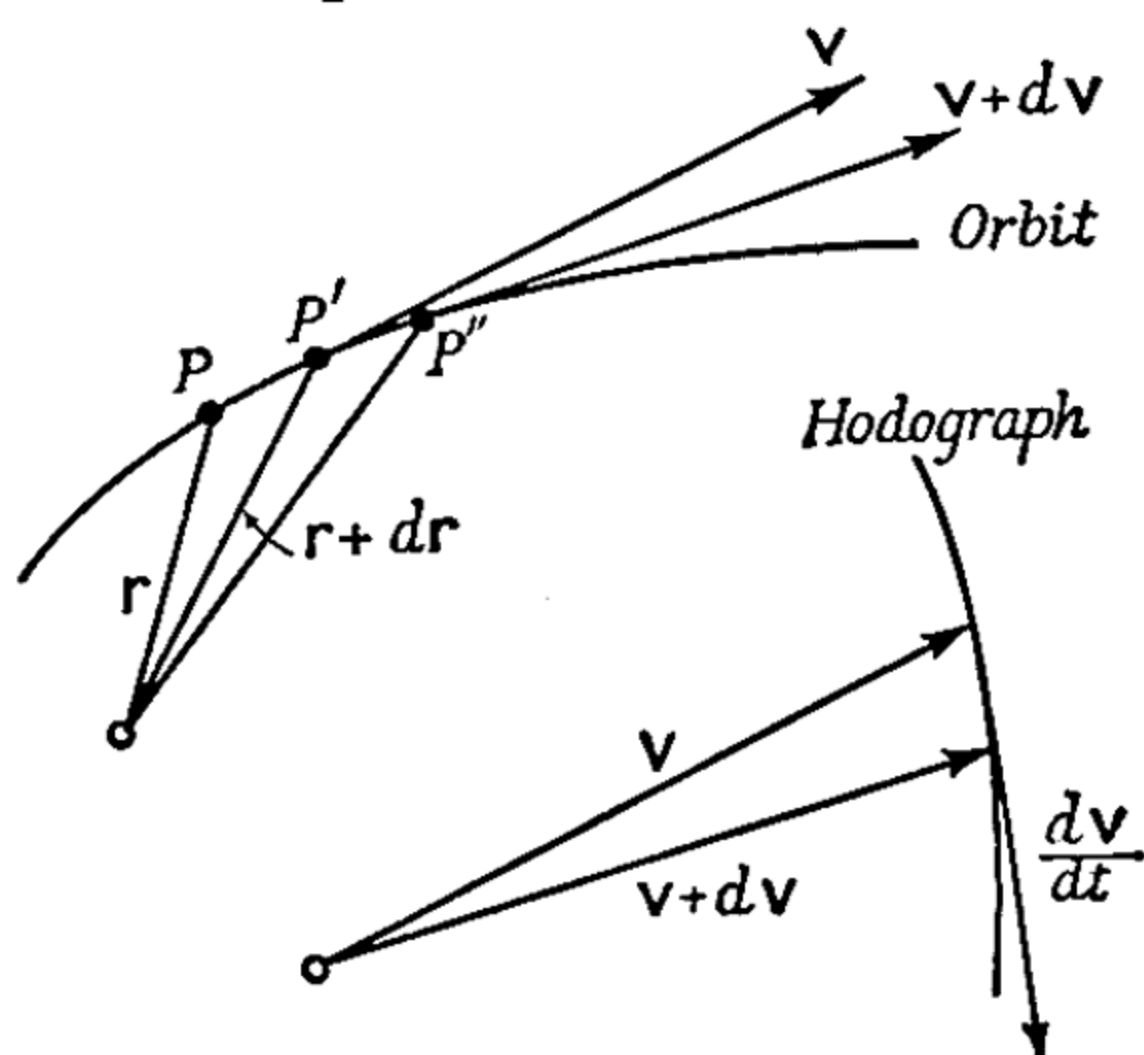


Fig. 1

is the vector  $\mathbf{v}$  resulting from taking the limit of  $d\mathbf{r}/dt$  for  $dt$  infinitely small, i.e.

$$\mathbf{v} = \frac{d\mathbf{r}}{dt} = \dot{\mathbf{r}}.* \quad . . . . . (1)$$

The direction of the velocity vector is thus given by the small vector  $d\mathbf{r}$  having the direction of the tangent to the path. Now, in general, the vector  $\mathbf{v}$  will itself be a function of the time. We obtain a vivid picture of the trend of the velocity if we draw the vectors  $\mathbf{v}$  from a new centre  $O_1$  as radius vectors. Then there is a definite terminus of a radius vector  $\mathbf{v}$  corresponding to every instant of time. If all these points are joined, we obtain a new curve, called the *hodograph* of the motion. While the particle  $P$  traverses the actual orbit, the "image point of the velocity", as defined above, moves over the hodograph.

\* When not explicitly written out, each differentiation with respect to time will be indicated by a dot over the appropriate quantity.

The *acceleration* is the limit of the change in  $\mathbf{v}$  per unit time, i.e.

$$\mathbf{a} = \frac{d\mathbf{v}}{dt} = \dot{\mathbf{v}} = \frac{d^2\mathbf{r}}{dt^2} = \ddot{\mathbf{r}}. \quad . . . . . (2)$$

Since  $\mathbf{v}$  is the radius vector of the point on the hodograph, the acceleration of the particle itself is equal to the velocity of its image point on the hodograph.

The velocity  $\mathbf{v}$  is said to be uniform only if the *vector*  $\mathbf{v}$  is constant. This definition requires that the motion be rectilinear. Motion on a curve in which the magnitude of the velocity is constant—i.e. where equal arcs are traversed in equal times—can nevertheless not be said to have uniform velocity.

We may resolve the acceleration in various ways:

(a) Tangential and normal acceleration:

$$\mathbf{v} = \frac{d\mathbf{r}}{dt} = \frac{d\mathbf{r}}{ds} \frac{ds}{dt} = v\mathbf{t}, \quad . . . . . (3)$$

where  $\mathbf{t}$  is the unit tangent vector (p. 18) and  $v$  is the magnitude of  $\mathbf{v}$ .

Differentiation gives

$$\mathbf{a} = \frac{dv}{dt} \mathbf{t} + v \frac{d\mathbf{t}}{ds} \frac{ds}{dt} = \frac{d^2s}{dt^2} \mathbf{t} + \frac{v^2}{\rho} \mathbf{n} \quad . . . . . (4)$$

[cf. eqn. (32), p. 19]. Here  $\mathbf{n}$  signifies the unit normal vector pointing toward the concave side of the curve, and  $\rho$  is the radius of curvature. The *tangential acceleration* has the magnitude

$$a_{\text{tan}} = \frac{dv}{dt} = \frac{d^2s}{dt^2}. \quad . . . . . (5)$$

The *normal acceleration*, directed toward the inside of the path, is

$$a_n = \frac{v^2}{\rho}. \quad . . . . . (6)$$

From (4) it is evident that the acceleration is always in the osculating plane.

(b) In rectangular co-ordinates:

$$\left. \begin{aligned} \mathbf{r} &= x\mathbf{i} + y\mathbf{j} + z\mathbf{k}, \\ \dot{\mathbf{r}} &= \dot{x}\mathbf{i} + \dot{y}\mathbf{j} + \dot{z}\mathbf{k}, \\ \ddot{\mathbf{r}} &= \ddot{x}\mathbf{i} + \ddot{y}\mathbf{j} + \ddot{z}\mathbf{k}. \end{aligned} \right\} . . . . . (7)$$

That is,

$$\left. \begin{aligned} v_x &= \dot{x}, & v_y &= \dot{y}, & v_z &= \dot{z} \\ a_x &= \ddot{x}, & a_y &= \ddot{y}, & a_z &= \ddot{z} \end{aligned} \right\} . . . . . (8)$$

(c) The resolution in plane polar co-ordinates  $r, \phi$  is also of importance. For this purpose we introduce two mutually perpendicular unit vectors  $e_r$  and  $e_\phi$ . These vectors have, respectively, the directions of increasing  $r$  and increasing  $\phi$ , and unlike the vectors  $i, j, k$ ,

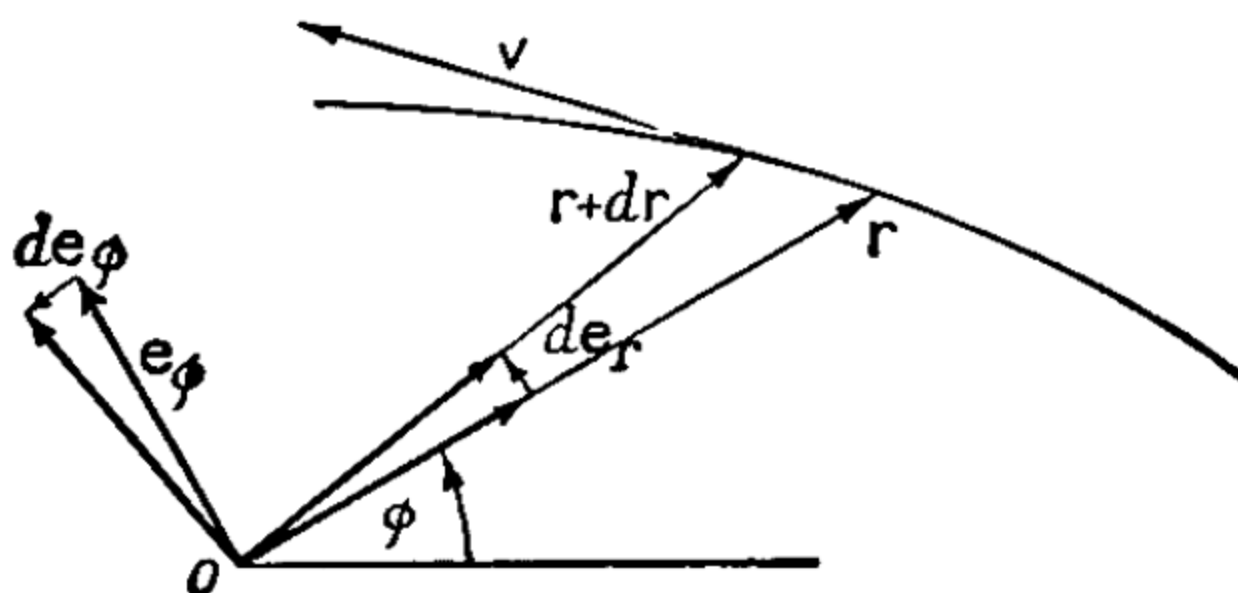


Fig. 2

change their direction with time. Since the derivative of a unit vector is perpendicular to the vector, we have (cf. fig. 2)

$$\frac{de_r}{dt} = e_\phi \frac{d\phi}{dt}, \quad \frac{de_\phi}{dt} = -e_r \frac{d\phi}{dt},$$

and we obtain, from  $r = re_r$ ,

$$\dot{r} = \dot{r}e_r + r\dot{e}_r = \dot{r}e_r + \dot{\phi}re_\phi,$$

and, by a second differentiation,

$$\begin{aligned} \ddot{r} &= \ddot{r}e_r + \dot{r}\dot{e}_r + \ddot{\phi}re_\phi + \dot{\phi}\dot{r}e_\phi + \dot{\phi}r\dot{e}_\phi \\ &= (\ddot{r} - r\dot{\phi}^2)e_r + (\ddot{\phi}r + 2\dot{\phi}\dot{r})e_\phi. \end{aligned}$$

Thus

$$\left. \begin{aligned} a_r &= \ddot{r} - r\dot{\phi}^2 \\ a_\phi &= \ddot{\phi}r + 2\dot{\phi}\dot{r} \end{aligned} \right\} \dots \dots \dots (9)$$

## 2. Newton's Second Law of Motion.

### (a) The law of inertia

The first of Newton's laws states that *every body tends to remain at rest or in uniform motion in a straight line, unless acted upon by external influences*. This law, commonly called the Law of Inertia, is by no means axiomatic, but represents an ingenious extrapolation of our physical experience; for we cannot free a body completely from external influences, since the gravitational effects of the earth and of the stars are always present.



(b) *Force and acceleration*

If we wish to alter the state of motion of a body, e.g. a sphere resting on a horizontal plane, by the use of our muscles, then a certain effort, increasing with the resulting change in velocity, is required. This effort, which we call *force*, is a direct sensory impression, and hence not further definable. The *direction* of the acceleration is determined by the direction in which we allow our muscles to work, so that force, like acceleration, is a vector. But from common experience we know that different amounts of force are necessary to produce a given acceleration, depending upon the nature of the body acted upon, e.g. a sphere of iron or of wood. Reciprocally, the effect of a given force depends upon the nature of the body to which it is applied. The direction of the acceleration, however, is the same for all bodies. We must therefore assign to each body a scalar property which we call the *inertial mass*  $m$ . The simplest equation embracing these factors is

$$\mathbf{F} = m\mathbf{a} = m \frac{d\mathbf{v}}{dt} = m \frac{d^2\mathbf{r}}{dt^2}. \quad \cdot \cdot \cdot \cdot (10)$$

Since we have introduced no factor of proportionality, the unit of force is determined by using the fundamental C.G.S. units, i.e. length in cm., mass in gm., and time in sec. The force which imparts to 1 gm. of mass an acceleration of 1 cm./sec.<sup>2</sup> is called 1 dyne.

Throughout Newtonian Mechanics, to which we limit ourselves for the present, it is tacitly assumed that the inertial mass  $m$  is a constant individual property of a body. A more cautious way of writing the fundamental law (10) leaves this question open:

$$\mathbf{F} = \frac{d}{dt}(m\mathbf{v}) = \frac{d\mathbf{u}}{dt}. \quad \cdot \cdot \cdot \cdot (10')$$

The quantity  $\mathbf{u} = m\mathbf{v}$  introduced here is called the *momentum*. Newton's second law, stated in terms of momentum, is then: *The force is equal to the change in momentum per second.\**

All about us in nature we see changes taking place in the state of motion of bodies, without exertion of muscular effort on our part. We interpret the cause of these changes as "forces" which operate on the bodies in the same way as our muscles do. The investigation of these forces, which may be of a very diverse nature, is an important problem in physics. In mechanics, however, we take the force to be

\* A more complete general statement reads: "Rate of change of momentum is proportional to the force producing it, and this change takes place in the direction in which the force acts."

given, and confine ourselves to computing the effects. If several forces act on a body, each produces its own acceleration:

$$\mathbf{F}_i = m \mathbf{a}_i.$$

But since the resultant acceleration is the vector sum of the separate accelerations, and the inertial mass  $m$  is the same for all,

$$m \mathbf{a} = m \sum \mathbf{a}_i = \sum \mathbf{F}_i = \mathbf{F}.$$

We may therefore replace all the separate applied forces  $\mathbf{F}_i$  by their resultant  $\mathbf{F}$  in order to calculate the acceleration.

Like most physical laws, the Newtonian Law of Motion has the form of a differential equation. The variation in any interval of time is obtained from this equation by continuously repeated application of the law, starting with given initial conditions, i.e. by integration of the differential equation. The unique power of theoretical physics is to be found in the circumstance that the laws of nature can be formulated in terms of differential equations. The fact that we can deduce descriptions of such apparently diverse phenomena as free fall, projectile motion and planetary motion from a single law depends on the circumstance that what happens in a short interval of time in any of these cases is deducible from the same relation.

In most cases eqn. (10) (p. 84) must be expressed in terms of the co-ordinates appropriate to the problem. Thus we obtain, in general, a system of three differential equations. Since the solution of three differential equations of the second order contains six constants, a complete description of the phenomenon is obtained only if these six constants are known. These are uniquely determined by giving the initial position and initial velocity, each having three components.

*Ex. 31.* A point moves subject to a periodic force  $\mathbf{F} = A \sin \omega t$ . Compute the motion for various initial velocities.

*Ex. 32.* The force acting on every mass  $m$  at the earth's surface is proportional to  $m$  and is directed toward the centre of the earth. In a limited region we can take the force to be constant, and the earth's surface to be plane. Integrate the equations of motion for this case, for various values of the initial velocity, and draw the orbit and the hodograph of the motion. (Theory of projectiles *in vacuo*.)

### 3. Time Integral and Path Integral of the Force. Work and Energy.

#### (a) *The time integral*

We can make two different statements for the effect of a force acting on a particle for a certain length of time, each of which leads to relationships having meaning. First, we can multiply the force by the time during which it acts; secondly, we can take the scalar product of the force by the vector displacement which it causes. If the force itself is variable, then instead of the finite time interval



we have the sum of the products of the instantaneous time values of the force by the time element, i.e. the integral  $\int \mathbf{F} dt$ . A similar formulation holds for the second product. If we apply the fundamental law, the time integral becomes

$$\int_{t_0}^t \mathbf{F} dt = \int_{t_0}^t m \frac{d^2 \mathbf{r}}{dt^2} dt = m \left( \frac{d\mathbf{r}}{dt} \right) - m \left( \frac{d\mathbf{r}}{dt} \right)_{t_0} = \mathbf{u} - \mathbf{u}_0 \quad (11)$$

*The time integral of the force (also called the impulse\* of the force) is equal to the change of momentum.*

The impulse law (11) is especially important for a large force which acts for a very short time  $\Delta t$ , so that the product  $\mathbf{F}\Delta t$  remains finite, and thus causes a finite change of momentum. Observation of the change of momentum gives the product  $\bar{\mathbf{F}}\Delta t$  immediately, without requiring a knowledge of the way the force varies. Here  $\bar{\mathbf{F}}$  represents a mean value of the force, viz. the time average.

### (b) The path integral of the force

If a force  $\mathbf{F}$  acts on a particle, the work done by  $\mathbf{F}$  when the particle moves through a displacement  $\dagger d\mathbf{r}$  is defined to be the scalar product  $\mathbf{F} d\mathbf{r}$ .

For a finite displacement along a curve  $C$ , the work is the sum of such contributions  $\mathbf{F} d\mathbf{r}$ , i.e. we have, for the work, the integral

$$W = \int_{r_1}^{r_2} \mathbf{F} d\mathbf{r}. \quad . . . . . (12)$$

If we substitute for  $\mathbf{F}$  the equal vector  $m(d^2\mathbf{r}/dt^2)$ , and introduce the time as variable of integration by using  $d\mathbf{r} = (d\mathbf{r}/dt)dt$ , then we obtain

$$\begin{aligned} W &= \int_{t_1}^{t_2} \mathbf{F} \frac{d\mathbf{r}}{dt} dt = \int_{t_1}^{t_2} m \frac{d^2 \mathbf{r}}{dt^2} \frac{d\mathbf{r}}{dt} dt \\ &= \int_{t_1}^{t_2} d \left[ \frac{1}{2} m \left( \frac{d\mathbf{r}}{dt} \right)^2 \right] = \frac{1}{2} m (v_2^2 - v_1^2). \quad . . . (13) \end{aligned}$$

The entire amount of work done by the force is thus given by the difference between the initial and final values of the expression  $\frac{1}{2}mv^2$ , which we call the *kinetic energy* of the particle. Thus *the total work done by the force is equal to the difference of the kinetic energy in the initial and final states*. From the definition of work, and the fundamental property of the scalar product, it follows that a force which is always perpendicular to the orbit of a particle can do no work, i.e.

\* In view of equation (11), the term "impulse" has come to be used not only for momentum but for the product  $m d\mathbf{r}/dt$ , although the word originally referred only to the left side of the equation.

† The path differential is called  $d\mathbf{r}$  instead of  $ds$  here to indicate explicitly that the position of the particle is given by the radius vector  $\mathbf{r}$ .



can cause no change in the kinetic energy. Such a force can change the direction of the velocity, but not its amount.

*Ex. 33.* A coin lies on top of a card which in turn rests on a drinking-glass. Why does the coin remain on the card if the latter is slowly drawn away, while if the card is snapped away suddenly, the coin falls into the glass? (Remember the empirical fact that the friction between solid bodies is independent of the relative velocity.)

*Ex. 34.* Decide whether or not the fatigue experienced in holding a weight with arm outstretched has any connexion with work done.

#### 4. Conservative Forces; Potential.

Consider a particle  $P$  in a field of force, i.e. in a portion of space at every point of which there exists a force  $\mathbf{F}$ , uniquely determined by the co-ordinates. Let the particle  $P$  be moved from  $P_0$  to  $P_1$  (fig. 3).

We divide forces into two classes, according as the work done in the displacement does not or does depend upon the path taken. This difference may be characterized somewhat differently if we consider also a return path  $P_1P_0$ . With the first kind of force, the integral along  $P_1P_0$  has the

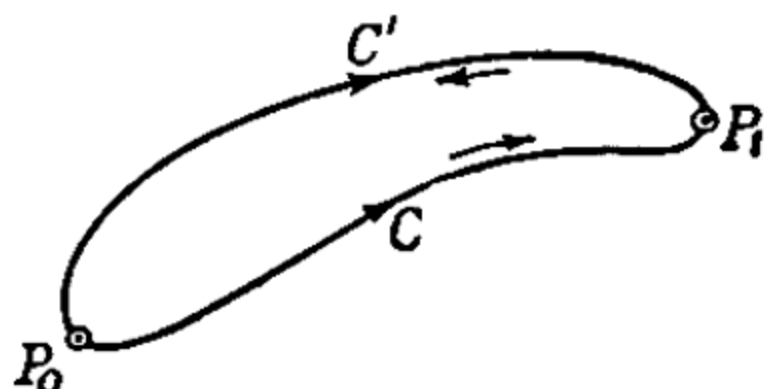


Fig. 3

opposite value, since the curve is traversed in the opposite sense. If, then, we traverse the closed curve  $P_0P_1P_0$ , the integral vanishes if the force is of the first kind. Forces of this type—to which, for example, the force of gravity belongs—are called *conservative forces*. If a body is projected vertically upward in a vacuum, with a given initial velocity, it will return to the starting-point with the same speed; the kinetic energy has not changed, and no work has been done on the complete path. The situation is quite different if the resistance of air or of any fluid operates. In this case the body returns to its point of projection with greatly reduced speed, and the integral  $\oint \mathbf{F} d\mathbf{r}$  cannot vanish, according to equation (13), p. 86.

Forces belonging to this second class are termed *non-conservative*. The names originate from the fact that, in the first case, the total mechanical energy is conserved—there is merely an interchange between kinetic and potential energy (cf. below)—while in the second instance, a transformation to other energy forms (principally heat) takes place.

When forces are conservative, their discussion can be simplified by introducing the “potential”. If the work done on a particle by the field of force during a displacement is independent of the path, then we can assign to every point of space the value of the work done in a displacement from a “standard position”  $r_1$  to the point  $r$ . Thus, besides the vector field of the force (determined by three num-

bers), we have a much simpler scalar field. For practical reasons we take the negative of the *work done by the field* of force and call it the *Potential*  $U$ . The change in  $U$  corresponding to a small displacement  $d\mathbf{r}$  is  $d\mathbf{r} \text{ grad } U$ . The work done is then

$$dW = \mathbf{F} d\mathbf{r} = -dU = -(\text{grad } U) d\mathbf{r}, \quad . \quad . \quad . \quad (14)$$

$$\text{i.e.} \quad \mathbf{F} = -\text{grad } U. \quad . \quad . \quad . \quad . \quad . \quad (14')$$

*Every conservative force may be represented as the negative of the gradient of a scalar point function, the potential.* For a finite displacement we have, by equation (13),

$$W = \int_{r_1}^{r_2} \mathbf{F} d\mathbf{r} = \frac{1}{2}m(v_2^2 - v_1^2) = - \int_{r_1}^{r_2} (\text{grad } U) d\mathbf{r} = U_1 - U_2. \quad (15)$$

Denoting the kinetic energy  $\frac{1}{2}mv^2$  by  $T$ , we have

$$T_2 + U_2 = T_1 + U_1 = \text{const.} \quad . \quad . \quad . \quad . \quad (16)$$

This is the Principle of the Conservation of Energy for mechanics.  $U$ , which has the dimensions of energy, is called the *Potential Energy*. Equation (16) thus means: *The sum of the kinetic and potential energies is constant for conservative forces.*

This energy principle represents a first integral of Newton's Second Law, since in (16) only  $v$  (i.e.  $ds/dt$ , the first derivative) appears. But since the Second Law, being a vector equation, is equivalent to three scalar differential equations, we need two additional first order equations in order to complete the first step in the integration of the equations of motion.

With a given field of force it is, as a rule, not difficult to write down the potential at once.  $\mathbf{F}$  may then be found by taking the gradient. Mathematically, this means the integration of the system

$$F_x = -\frac{\partial U}{\partial x}, \quad F_y = -\frac{\partial U}{\partial y}, \quad F_z = -\frac{\partial U}{\partial z},$$

to which one must have recourse in case of doubt.

*Ex. 35.* A particle is bound to a fixed centre by an elastic force, i.e. one proportional to the distance from the centre. What is the potential corresponding to this force?

## 5. Central Forces; the Law of Areas.

A central force is one whose representative vector always points toward a fixed point  $O$ . The magnitude of the force may be an arbitrary point function. Naturally, we choose the point  $O$  as origin, so that  $\mathbf{F} = f(x, y, z)\mathbf{r}$ .

In the case of central forces, an important theorem may be derived immediately. This law furnishes two additional first integrals of the fundamental equations. We form the vector product of  $\mathbf{r}$  with each side of the equation representing Newton's Second Law. Since  $\mathbf{F}$  is in the direction of  $\mathbf{r}$ , and since the vector product of two parallel vectors is zero, we have

$$0 = m \left[ \mathbf{r} \frac{d^2 \mathbf{r}}{dt^2} \right] \quad \text{or} \quad \left[ \mathbf{r} \frac{d^2 \mathbf{r}}{dt^2} \right] = 0.$$

The integral may be given at once:

$$\left[ \mathbf{r} \frac{d\mathbf{r}}{dt} \right] = 2\mathbf{c}, \quad . . . . . (17)$$

for differentiation gives

$$\left[ \frac{d\mathbf{r}}{dt} \frac{d\mathbf{r}}{dt} \right] + \left[ \mathbf{r} \frac{d^2 \mathbf{r}}{dt^2} \right] = \left[ \mathbf{r} \frac{d^2 \mathbf{r}}{dt^2} \right] = 0.$$

The vector product  $[\mathbf{r} d\mathbf{r}/dt]$  has a simple geometric significance. The magnitude of  $[\mathbf{r} d\mathbf{r}]$  is equal to twice the area swept out by the radius vector in the time  $dt$  (fig. 4). If we divide the area for  $dt$  by  $dt$  we obtain the area swept over by the radius vector in unit time—the so-called areal velocity  $\mathbf{c}$ . Equation (17) thus states that *the areal velocity is constant for motion under a central force*. This is *Kepler's Second Law of Planetary Motion*. But equation (17) says more. If the product  $[\mathbf{r}(d\mathbf{r}/dt)]$  is equal to a constant vector, the normal to the plane determined by  $\mathbf{r}$  and  $d\mathbf{r}$  has a fixed direction, i.e. *the orbit is plane*. For this reason equation (17) represents only two scalar equations which, however, together with the energy equation, are the complete first integral of the equations of motion.

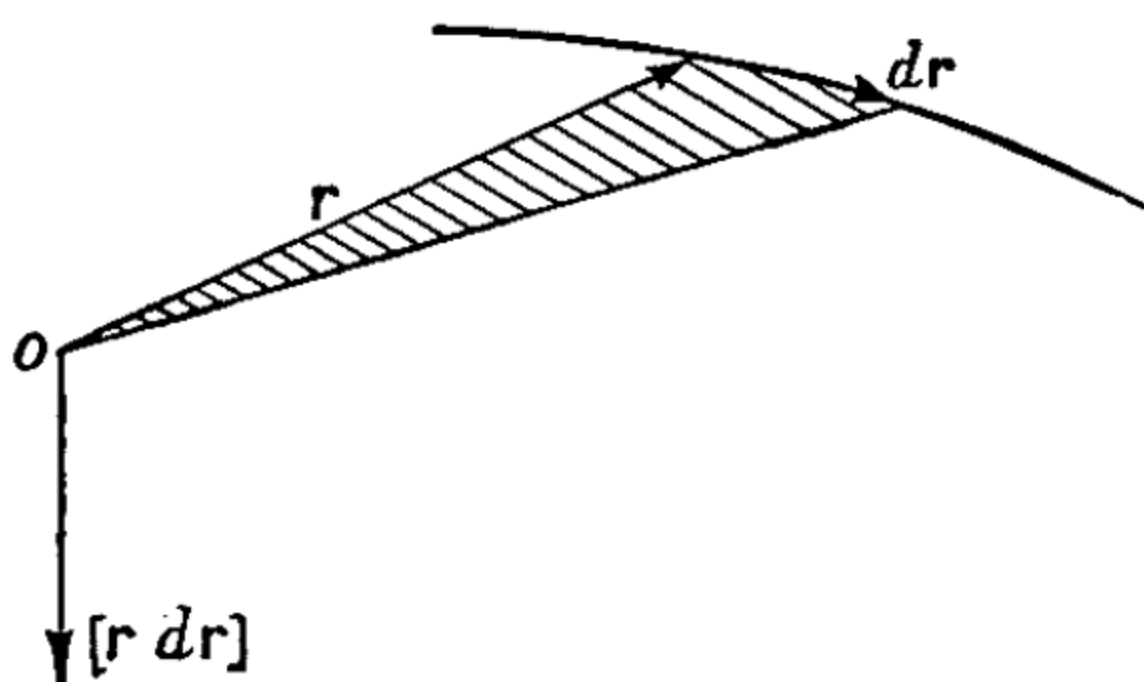


Fig. 4

The vector  $m \left[ \mathbf{r} \frac{d\mathbf{r}}{dt} \right]$  is usually called the *moment of momentum* or *angular momentum*  $\mathbf{P}$ , about  $O$ .

**Ex. 36.** Verify the following generalization of the Law of Areas: if the force acting on a particle always passes through a certain fixed axis, the Law of Areas holds for the projection of the orbit on a plane perpendicular to the axis.

**Ex. 37.** From the equation of the orbit  $r = f(\phi)$  and the areal velocity  $c$  deduce the Law of Force. Use equation (9). Application of this line of thought to the empirically known Kepler ellipses led Newton to the discovery of the Law of Gravitation.



## 6. Gravitational Forces; Planetary Motion.

Starting with Kepler's Laws, at that time known as empirical facts, Newton derived his general law of attraction, specifying the forces which determine the motions of the heavenly bodies. The Newtonian Law of Universal Gravitation states that there is a force of attraction between any two masses  $m_1$  and  $m_2$  which is directly proportional to the product of the masses and inversely proportional to the square of their distance apart. That is

$$F = \gamma \frac{m_1 m_2}{r_{12}^2}. \quad . . . . . (18)$$

Up to this point we have encountered the mass of a body only in connexion with inertia; here it appears in an entirely different form—as gravitational mass (weight). It does not follow that inertial and gravitational mass are identical; i.e. more accurately stated, that the ratio of inertial to gravitational mass is a constant (contained in  $\gamma$ ) for all bodies. If we assume the two kinds of mass to be the same\* the constant  $\gamma$  has the value  $6.66 \times 10^{-8}$  C.G.S. units.

If we neglect small perturbations, the gravitational field of the sun alone determines the orbits of bodies in the solar system. On account of its predominantly large mass  $M$ , we can consider the sun fixed at the origin of a co-ordinate system. Strictly speaking, we should investigate the motion about the common centre of gravity, which is at rest (cf. Chap. VI). The force acting upon the body of mass  $m$  is

$$F = -\gamma \frac{Mm}{r^3} \mathbf{r}. \quad . . . . . (19)$$

The corresponding potential is

$$U = -\gamma \frac{Mm}{r}, \quad . . . . . (20)$$

as may be verified by forming the gradient.

Since the force is central, the orbit is plane. We can describe it most simply in plane polar co-ordinates with the sun at the centre. The first integrals of the equation of motion, the Energy Principle and the Law of Areas, may be written down at once:

$$\frac{1}{2} mv^2 - \gamma \frac{Mm}{r} = \frac{1}{2} mv_0^2 - \gamma \frac{Mm}{r_0} \quad (\text{Conservation of Energy}),$$

$$\left| \left[ r \frac{d\mathbf{r}}{dt} \right] \right| = r^2 \phi = 2c \quad (\text{Law of Areas}).$$

\* The equality of gravitational and inertial mass is a fundamental postulate of the Generalized Theory of Relativity. All experimental attempts to detect individual differences with various substances have been unsuccessful.

If we wish to describe the course of events in time, we must determine  $r$  and  $\phi$  as functions of  $t$ , by integrating again. Instead of doing this, however, we limit ourselves to the derivation of the equation of the orbit described by the body. For this purpose we eliminate the time from the two equations. Using

$$v^2 = \dot{r}^2 + r^2 \dot{\phi}^2 \text{ (cf. p. 83), } \frac{d}{dt} = \dot{\phi} \frac{d}{d\phi}, \quad \dot{\phi} = \frac{2c}{r^2},$$

this is readily accomplished. From the energy equation,

$$\frac{4c^2}{r^4} \left( \frac{dr}{d\phi} \right)^2 + \frac{4c^2}{r^2} - \frac{2\gamma M}{r} = v_0^2 - \frac{2\gamma M}{r_0}.$$

The variables  $r$  and  $\phi$  are immediately separable and we have

$$d\phi = \frac{2c \frac{dr}{r^2}}{\sqrt{\left( v_0^2 - \frac{2\gamma M}{r_0} \right) + \frac{2\gamma M}{r} - \frac{4c^2}{r^2}}}.$$

If we introduce a new variable  $u = 1/r$ , we obtain

$$\phi + C = - \int \frac{2c du}{\sqrt{\left( v_0^2 - \frac{2\gamma M}{r_0} \right) + 2\gamma M u - 4c^2 u^2}}.$$

This integral is of the familiar arc-cosine form

$$- \int \frac{dx}{\sqrt{a + 2bx - hx^2}} = \frac{1}{\sqrt{h}} \cos^{-1} \left( \frac{b - hx}{\sqrt{b^2 + ah}} \right).$$

If we introduce the abbreviation

$$a = \sqrt{v_0^2 - \frac{2\gamma M}{r_0} + \frac{\gamma^2 M^2}{4c^2}},$$

and take the cosine of each member, we obtain

$$r = \frac{\frac{4c^2}{\gamma M}}{1 + \frac{2ca}{\gamma M} \cos(\phi + C)} = \frac{p}{1 + \epsilon \cos(\phi + C)}. \quad (21)$$

We may measure  $\phi$  from the maximum  $r$ , so that  $C = \pi$ . This gives the familiar polar equation of a conic with origin at one focus. This is *Kepler's First Law*, which states that *the orbits of bodies in the solar system are conics, with the sun occupying one focus*.

The orbit is an ellipse, parabola or hyperbola, according as the numerical value of the eccentricity  $\epsilon$  is less than, equal to, or

greater than unity. Putting in the values of  $\epsilon$  and  $a$ , we have an ellipse if

$$\frac{2c}{\gamma M} \sqrt{v_0^2 - \frac{2\gamma M}{r_0} + \frac{\gamma^2 M^2}{4c^2}} < 1$$

or

$$\frac{v_0^2}{2} < \frac{\gamma M}{r_0} \quad \text{or} \quad \frac{1}{2} m v_0^2 < \frac{\gamma M m}{r_0};$$

a parabola if

$$\frac{1}{2} m v_0^2 = \frac{\gamma M m}{r_0};$$

and a hyperbola if

$$\frac{1}{2} m v_0^2 > \frac{\gamma M m}{r_0}.$$

That is, the orbit is an ellipse, parabola or hyperbola according as, initially, the kinetic energy is less than, equal to, or greater than the potential energy with sign changed. Thus, for an elliptic orbit, the total energy is negative.

The following additional relations may be derived for the case of an ellipse. The parameter of the ellipse

$$p = \frac{b^2}{a} = a(1 - \epsilon^2) \quad \begin{pmatrix} a = \text{semi-major axis} \\ b = \text{semi-minor axis} \end{pmatrix}$$

is equal to  $4c^2/\gamma M$ , by equation (21). Hence

$$1 - \epsilon^2 = \frac{4c^2}{\gamma M a}.$$

If we denote the period by  $T$ , the area of the ellipse is

$$cT = \pi ab = \pi a^2 \sqrt{1 - \epsilon^2} = \frac{2\pi a^2 c}{\sqrt{\gamma M a}}.$$

Thus

$$T^2 = \frac{4\pi^2 a^3}{\gamma M}. \quad \dots \dots \dots (22)$$

This is Kepler's Third Law: *The squares of the periods of the planets are proportional to the cubes of their orbital major axes.*

If instead of using the parameter  $p$  to compute  $1 - \epsilon^2$  we use the value of the eccentricity itself from (21), we have, on setting the two equal:

$$\frac{4c^2}{\gamma M a} = 1 - \frac{4c^2}{\gamma^2 M^2} \left( v_0^2 - \frac{2\gamma M}{r_0} + \frac{\gamma^2 M^2}{4c^2} \right)$$

or

$$\frac{1}{2} m v_0^2 - \frac{\gamma M m}{r_0} = E = - \frac{\gamma M m}{2a}. \quad \dots \dots (23)$$

This means that the total energy in an elliptical orbit is a function only of the size of the major axis, and all ellipses having the same major axis

(including the circle) have the same energy. This proposition simplifies many computations in the older atomic mechanics. The total energy has a negative sign because of the dominating magnitude of the potential energy, which was so specified upon introduction of the potential that it vanishes at infinity but is negative at finite distances.

## 7. Quasi-elastic Forces and Harmonic Vibrations.

If a body is bound to a fixed point by means of a spring or a rubber band, any small displacement from the position of rest calls forth a restoring force proportional to the displacement and directed toward the position of rest, which we take as origin. That is,

$$\mathbf{F} = -k\mathbf{r}.$$

We shall call all central forces of this form *quasi-elastic forces*, even though, as may well happen, the force is not of an elastic nature (cf. Ex. 38, p. 94). The differential equation becomes

$$m \frac{d^2 \mathbf{r}}{dt^2} + k\mathbf{r} = 0. \quad . \quad . \quad . \quad . \quad . \quad (24)$$

We might now, as before, use the Principle of Energy and the Law of Areas to get first integrals, but direct integration, twice repeated, is just as simple here. If we try the solution

$$\mathbf{r} = a e^{\lambda t}, \quad \dot{\mathbf{r}} = \lambda^2 a e^{\lambda t},$$

we obtain for  $\lambda$  the "characteristic equation"

$$m\lambda^2 + k = 0,$$

whose roots are

$$\lambda = i \sqrt{\frac{k}{m}} \quad \text{and} \quad \lambda = -i \sqrt{\frac{k}{m}}.$$

The general solution, containing two constant vectors, is

$$\mathbf{r} = a e^{i \sqrt{\frac{k}{m}} t} + b e^{-i \sqrt{\frac{k}{m}} t}. \quad . \quad . \quad . \quad . \quad (25)$$

Using real functions (cf. p. 47), this is

$$\mathbf{r} = \mathbf{A} \cos \sqrt{\frac{k}{m}} t + \mathbf{B} \sin \sqrt{\frac{k}{m}} t. \quad . \quad . \quad . \quad (25')$$

The meaning of the vectors  $\mathbf{A}$  and  $\mathbf{B}$  is readily ascertained. For

$$t = 0,$$

$$\mathbf{r}_0 = \mathbf{A},$$



$$\left(\frac{dr}{dt}\right)_0 = v_0 = B \sqrt{\frac{k}{m}},$$

$$\text{i.e.} \quad \mathbf{r} = r_0 \cos \sqrt{\frac{k}{m}} t + v_0 \sqrt{\frac{m}{k}} \sin \sqrt{\frac{k}{m}} t. \quad \cdot \cdot (25'')$$

The vector  $\mathbf{r}$  is thus the resultant of two vectors which have the directions of the initial radius vector and initial velocity, respectively, and whose magnitudes are simple periodic functions of the time. To obtain the orbit, we introduce the unit vectors  $\mathbf{e}_r$  and  $\mathbf{e}_v$ . In this oblique co-ordinate system, the co-ordinates of  $P$  are

$$\xi = r_0 \cos \sqrt{\frac{k}{m}} t, \quad \eta = v_0 \sqrt{\frac{m}{k}} \sin \sqrt{\frac{k}{m}} t.$$

$$\text{Elimination of } t \text{ gives} \quad \frac{\xi^2}{r_0^2} + \frac{\eta^2}{v_0^2 m/k} = 1.$$

This is the equation of an ellipse in oblique co-ordinates, the co-ordinate axes being in the directions of two conjugate diameters. *The general orbit of a particle moving under the influence of a quasi-elastic force is the ellipse having the initial position vector and the initial velocity vector as conjugate radii, and having its centre at the centre of force.* If  $v_0$  has the direction of  $\mathbf{r}_0$ , the vibration is linear. According to equation (25), the frequency of a harmonic vibration is given by  $1/2\pi$  times the square root of the quotient of the elastic constant by the mass:

$$\nu = \frac{1}{2\pi} \sqrt{\frac{k}{m}}. \quad \cdot \cdot \cdot \cdot \cdot (26)$$

*Ex. 38.* A particle constrained to move on a straight line is subjected to an electrostatic force of attraction  $F_1 = -\alpha/x^2$  and a repulsive force  $F_2 = +\beta/x^{10}$ . Calculate the frequency of the vibration resulting from a small displacement from the position of equilibrium  $x_0$ .

## 8. Harmonic Vibrations with Frictional Resistance.

In addition to the quasi-elastic force, let the particle be subjected to a frictional force, as is actually the case with vibratory motion in practice. Let us assume that this resisting force is proportional to the speed of the particle. A resistance of this kind exists when the particle moves at low speeds in the air or in any other fluid. Friction between solid bodies is, on the contrary, independent of the speed, within wide limits. Whatever the law of dependence of the resisting force on the velocity may be, this force is always opposite in direction to the velocity of the particle. The total force acting on  $P$  is therefore

$$F = -kr - \beta \frac{dr}{dt}.$$

We limit ourselves, in what follows, to rectilinear motion, and call the co-ordinate  $x$ . Then the equation of motion is

$$m \frac{d^2x}{dt^2} + \beta \frac{dx}{dt} + kx = 0. \quad \dots \dots (27)$$

To integrate, we again try

$$x = ae^{\lambda t},$$

which leads to the "characteristic equation"

$$m\lambda^2 + \beta\lambda + k = 0.$$

Its roots are

$$\lambda_1 = -\frac{\beta}{2m} + \sqrt{\frac{\beta^2}{4m^2} - \frac{k}{m}} \text{ and } \lambda_2 = -\frac{\beta}{2m} - \sqrt{\frac{\beta^2}{4m^2} - \frac{k}{m}}.$$

Two cases must be distinguished here:

(a)  $\beta^2 \geq 4km$ . In this case only real quantities appear, and the general solution for  $\beta^2 > 4km$  is

$$x = e^{-\frac{\beta}{2m}t} \left( ae^{\sqrt{\frac{\beta^2}{4m^2} - \frac{k}{m}}t} + be^{-\sqrt{\frac{\beta^2}{4m^2} - \frac{k}{m}}t} \right). \quad \dots \dots (28)$$

In this *aperiodic* case, corresponding to large damping, the particle gradually returns to its equilibrium position, which it reaches at  $t = \infty$ .

Even the limiting case  $\beta^2 = 4km$  results in an aperiodic solution. Superficially, it would seem that a double root of the characteristic equation would yield only one constant of integration. However, it is seen that here  $bte^{\lambda t}$  is also a solution, so that the general solution is

$$x = e^{-\frac{\beta}{2m}t} (a + bt) = e^{-\sqrt{\frac{k}{m}}t} (a + bt). \quad \dots \dots (29)$$

(b)  $\beta^2 < 4km$ . There are now imaginary exponents:

$$x = e^{-\frac{\beta}{2m}t} \left( ae^{i\sqrt{\frac{k}{m} - \frac{\beta^2}{4m^2}}t} + be^{-i\sqrt{\frac{k}{m} - \frac{\beta^2}{4m^2}}t} \right).$$

Applying the Euler formula, and introducing new constants of integration, we obtain the real form

$$x = e^{-\frac{\beta}{2m}t} \left( A \cos \sqrt{\frac{k}{m} - \frac{\beta^2}{4m^2}}t + B \sin \sqrt{\frac{k}{m} - \frac{\beta^2}{4m^2}}t \right),$$

or 
$$x = Ce^{-\frac{\beta}{2m}t} \cos \left( \sqrt{\frac{k}{m} - \frac{\beta^2}{4m^2}}t - \gamma \right). \quad \dots \dots (30)$$

We again have a harmonic vibration whose amplitude, however, decreases exponentially (cf. fig. 5). This damped vibration has the following important properties: The frequency

$$\nu = \frac{1}{2\pi} \sqrt{\frac{k}{m} - \frac{\beta^2}{4m^2}}$$

is less than

$$\nu_0 = \frac{1}{2\pi} \sqrt{\frac{k}{m}},$$

that of the undamped vibration. The ratio of two successive maximum deflections in the same direction is

$$p = e^{\frac{\beta T}{2m}}. \quad . . . . . (31)$$

The natural logarithm of this ratio is called the *logarithmic decrement*  $\delta$ . If the damping is small, we make no serious error by putting the period of the undamped motion in place of the period  $T$ . Thus is obtained the simple expression

$$\delta = \frac{\beta}{2m} T \approx \frac{\pi\beta}{\sqrt{mk}}. \quad . . . . . (31')$$

*Note:* In mechanics, especially in treating the problem of the pendulum, the period is often taken to be half the quantity here designated by  $T$ . It is thus taken as the time between an extreme deflection on one side and the next one on the opposite side. Correspondingly, the logarithmic decrement is often referred to the half-period.

## 9. Forced Vibrations; Resonance.

In addition to the forces previously taken to operate during a harmonic vibration, let us assume that a periodic external force is present. This force may have the form

$$F_a = F_0 \cos \omega t.$$

With these three forces—quasi-elastic, frictional, and applied—the equation of motion becomes

$$m \frac{d^2x}{dt^2} + \beta \frac{dx}{dt} + kx = F_0 \cos \omega t, \quad . . . (32)$$

Introducing the characteristic angular frequency of the free, undamped vibration,  $\omega_0^2 = k/m$ , we have

$$\frac{d^2x}{dt^2} + \frac{\beta}{m} \frac{dx}{dt} + \omega_0^2 x = \frac{F_0}{m} \cos \omega t. \quad . . . (32')$$

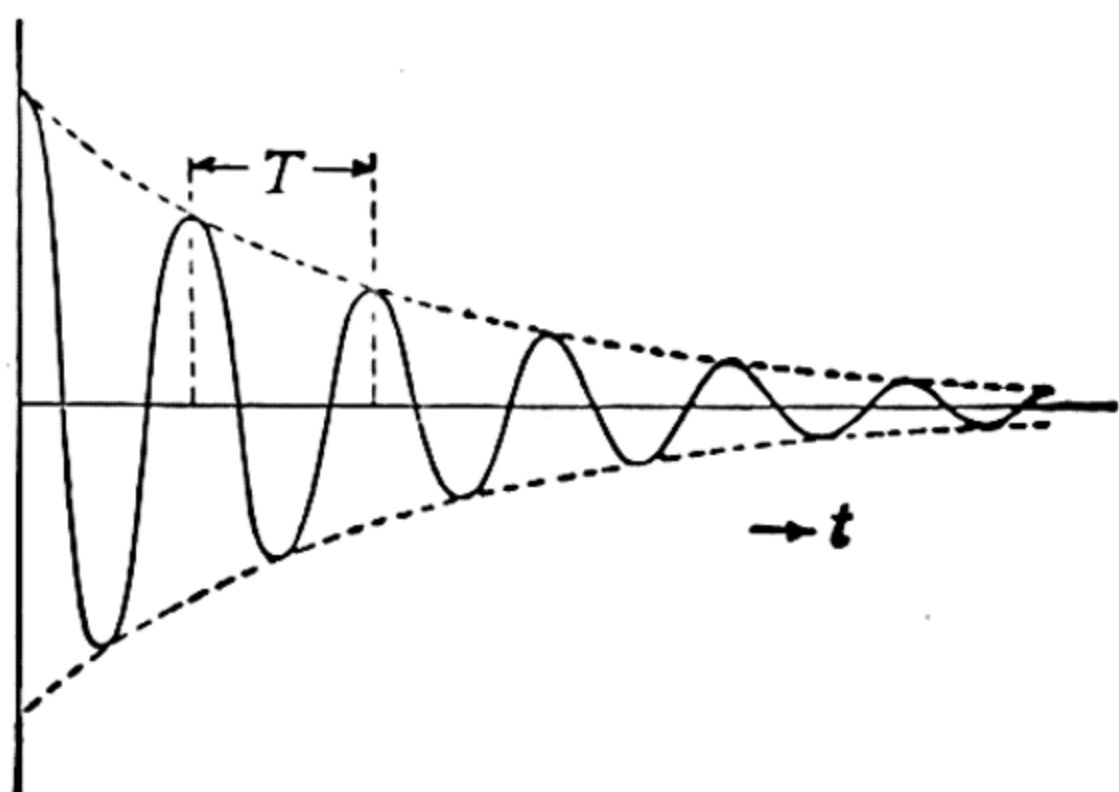


Fig 5



To integrate this equation we use a theorem from the theory of linear differential equations, which states: If  $g(t)$  is a particular integral of this non-homogeneous equation, and if  $f(t, A, B)$  is the general integral of the corresponding homogeneous equation,

$$\frac{d^2x}{dt^2} + \frac{\beta}{m} \frac{dx}{dt} + \omega_0^2 x = 0,$$

then the sum  $g(t) + f(t, A, B)$  is the general integral of the non-homogeneous equation. On account of the linearity of the differential equation (32')—i.e. the absence of products of the dependent variable and its derivatives—the sum  $g(t) + f(t, A, B)$  satisfies the equation, as may be seen by substitution. Since this solution contains two constants of integration, it is also the general solution. The integral of the homogeneous equation is known from the preceding paragraph, and so we need only find a particular integral of the non-homogeneous equation. Since the force  $F_a$  is a periodic function of angular frequency  $\omega$ , the trial of a periodic function of this kind is suggested. But since the first as well as the second derivative of  $x$  appears, it is evident that a sine or a cosine function *alone* cannot satisfy the equation. We thus try

$$x = p \cos \omega t + q \sin \omega t = \sqrt{p^2 + q^2} \cos(\omega t - \phi)$$

where the coefficients  $p$  and  $q$  are still undetermined. If we substitute, in equation (32'), the condition that the coefficients of  $\cos \omega t$  and of  $\sin \omega t$  must vanish leads to the following equations for  $p$  and  $q$ :

$$p(\omega_0^2 - \omega^2) + q \frac{\beta \omega}{m} = \frac{F_0}{m},$$

$$\frac{\beta \omega}{m} p - q(\omega_0^2 - \omega^2) = 0.$$

From these we find

$$\left. \begin{aligned} p &= \frac{m F_0 (\omega_0^2 - \omega^2)}{m^2 (\omega_0^2 - \omega^2)^2 + \beta^2 \omega^2}, & q &= \frac{\beta \omega F_0}{m^2 (\omega_0^2 - \omega^2)^2 + \beta^2 \omega^2} \\ \sqrt{p^2 + q^2} &= \frac{F_0}{\sqrt{m^2 (\omega_0^2 - \omega^2)^2 + \beta^2 \omega^2}}, & \tan \phi &= \frac{q}{p} = \frac{\beta \omega}{m (\omega_0^2 - \omega^2)}. \end{aligned} \right\} \quad (33)$$

The general solution for small damping [Case (b), p. 95] is

$$\begin{aligned} x &= C e^{-\frac{\beta}{2m} t} \cos \left( \sqrt{\frac{k}{m} - \frac{\beta^2}{4m^2}} t - \gamma \right) \\ &\quad + \frac{F_0}{\sqrt{m^2 (\omega_0^2 - \omega^2)^2 + \beta^2 \omega^2}} \cos(\omega t - \phi). \end{aligned} \quad (34)$$



The first part represents a damped vibration having the period of the characteristic free vibration of the damped system, i.e. it corresponds to a transient phenomenon determined by the initial conditions, which determine  $C$  and  $\gamma$ . This part dies out after some time. The second part represents a permanent vibration of the same period as the exciting force. This is the steady state. For stronger damping [Case (a)] the first part becomes aperiodic, while the form of the second part remains unchanged.

We could have found the particular integral which describes the steady state much more simply by using the vector diagram (p. 52) which always leads to a result rapidly in problems involving the determination of the permanent modes, or stationary states, in vibration processes. If we write for the exciting force,  $F_a = F_0 e^{i\omega t}$ , and take the solution to be of the form

$$x = ae^{i(\omega t - \phi)},$$

then, by equation (32') (p. 96) we must have

$$a \left( -\omega^2 + \frac{\beta}{m} i\omega + \omega_0^2 \right) e^{i(\omega t - \phi)} = \frac{F_0}{m} e^{i\omega t},$$

or 
$$a \left( \omega_0^2 - \omega^2 + i \frac{\beta\omega}{m} \right) = \frac{F_0}{m} e^{i\phi}; \quad . . . (33')$$

the addition of the complex numbers, or vectors, on the left (fig. 6) gives a sum of modulus

$$a \sqrt{(\omega_0^2 - \omega^2)^2 + \frac{\beta^2 \omega^2}{m^2}}$$

from the vector diagram (fig. 6). This has to be equal to  $F_0/m$ , i.e.

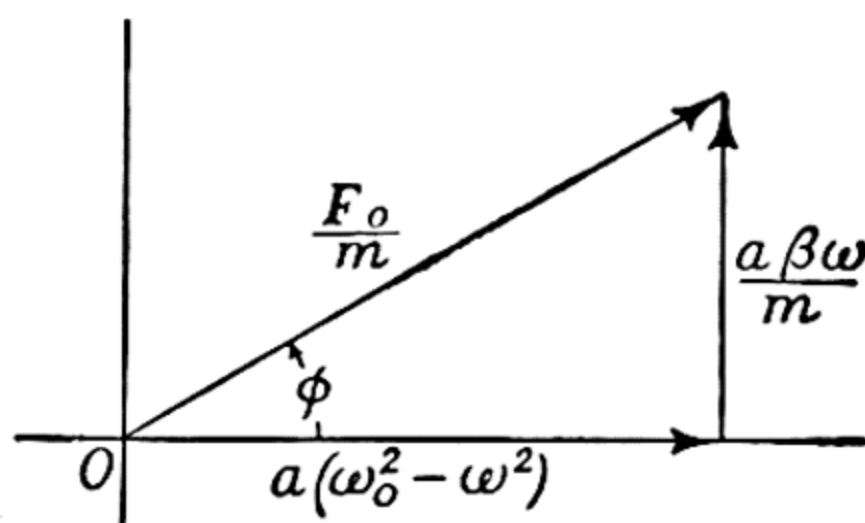


Fig. 6

$$a = \frac{F_0}{\sqrt{m^2(\omega_0^2 - \omega^2)^2 + \beta^2 \omega^2}}. \quad . . . (33'')$$

The two complex members (vectors) of equation (33') must agree in direction as well as in magnitude, i.e.

$$\tan \phi = \frac{\beta\omega}{m(\omega_0^2 - \omega^2)}. \quad . . . (33''')$$

Thus the difference in phase between the exciting force and the forced vibration, as well as the amplitude of the latter, both depend upon  $\omega_0^2 - \omega^2$ . The phase difference approaches zero for the case of very slow vibrations, and increases to  $\pi/2$  for  $\omega = \omega_0$  (resonance; see below), becoming  $\pi$  for infinitely rapid oscillations. When  $\omega$  is given, the amplitude is greatest when  $\omega_0 = \omega$ ; i.e. when the frequency of the

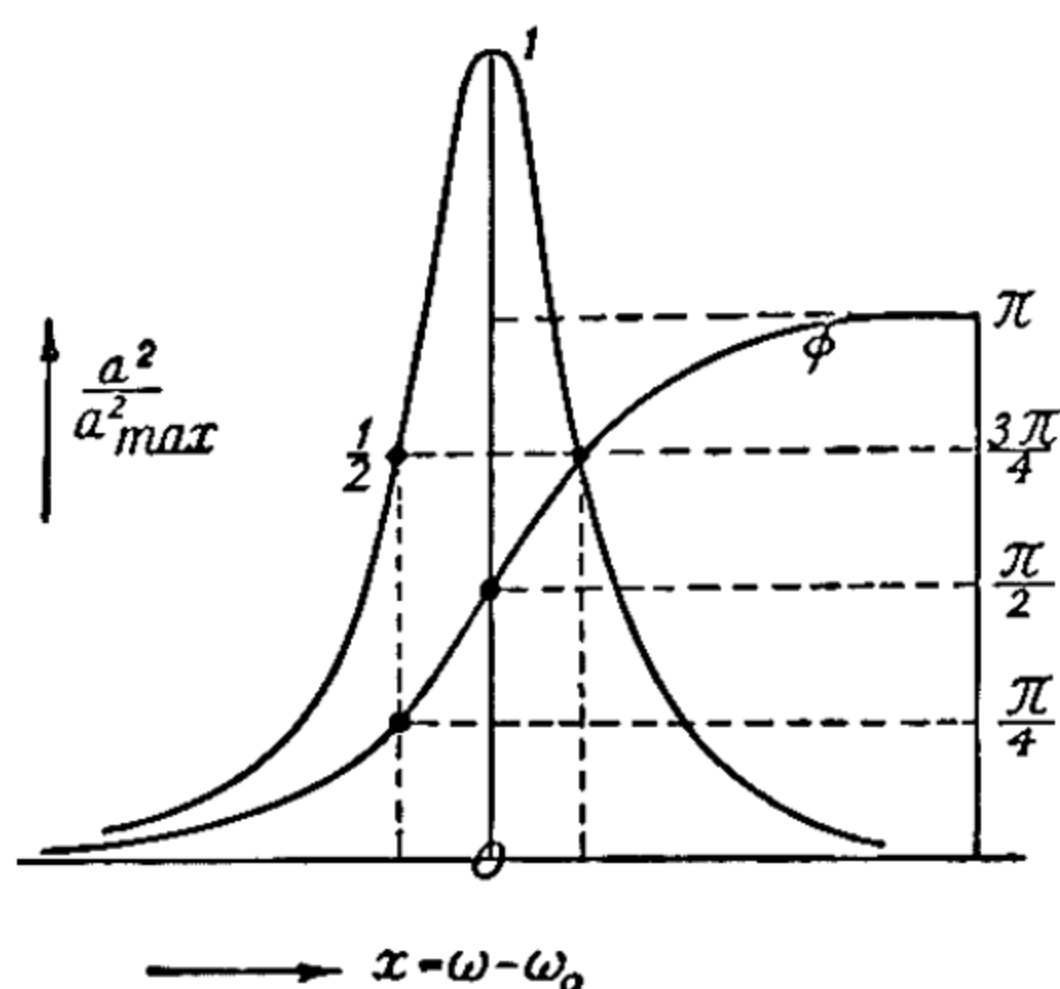


Fig. 7

exciting force coincides with the natural frequency of the *undamped* system, *not* with the natural frequency of the actual system. In this case we speak of *resonance*.\* In the absence of all damping the amplitude becomes infinite at resonance. To show how the amplitude depends on the frequency of the exciting force, i.e. to draw the so-called resonance curve, it is convenient to choose the square of the amplitude as ordinate. If we use the value for

$a^2$  from equation (33''), it follows that

$$a^2 = \frac{F_0^2}{m^2(\omega_0^2 - \omega^2)^2 + \beta^2\omega^2} = \frac{\beta^2 a_{\max}^2}{m^2 \frac{(\omega_0^2 - \omega^2)^2}{\omega^2} + \beta^2}.$$

Introducing the "discord"  $\omega - \omega_0$  as independent variable  $x$  and putting, as an approximation,

$$\omega_0 + \omega = 2\omega, \quad \omega_0^2 - \omega^2 = -2\omega x$$

in the neighbourhood of the resonance point, we obtain this simple equation for the resonance curve (fig. 7):

$$\frac{a^2}{a_{\max}^2} = \frac{\beta^2}{4m^2 x^2 + \beta^2}. \quad \dots \quad (34)$$

For  $\frac{a^2}{a_{\max}^2} = \frac{1}{2} = \frac{\beta^2}{4m^2 \bar{x}^2 + \beta^2}$ ,  $\bar{x}$  becomes  $\frac{\beta}{2m}$ ,

or  $\omega - \omega_0 = \frac{\beta}{2m} = \frac{\delta}{T} = \frac{\delta\omega_0}{2\pi}$  (cf. equation (31)),

or  $\frac{\bar{\omega} - \omega_0}{\omega_0} = \frac{\delta}{2\pi}. \quad \dots \quad (35)$

We therefore have the theorem: *Half the interval between the two frequencies for which the square of the amplitude has half its maximum value, these frequencies being measured in terms of the free undamped*

\* Strictly, this is the *velocity resonance* (cf. Ex. 39).

frequency as unit, is approximately equal to the logarithmic decrement divided by  $2\pi$ . That is, the weaker the damping of the system, the sharper the resonance. If we introduce the same simplification into (33'''), we see that the phase differences at the positions of half-value are  $\pi/4$  and  $3\pi/4$  respectively.

*Ex. 39.* If the equation of the resonance curve is calculated without resorting to the approximations made in the text, it is found that the maximum of the square of the amplitude does not occur at the natural frequency of the damped motion, nor at that of the undamped vibration. It is true, however, that the particle has its maximum energy (defined as the kinetic energy while passing through the zero-point) at  $\omega = \omega_0$ . The maximum of the time average of the energy comes at yet another place. Compute the positions of the maxima of the specified quantities.

## 10. Non-harmonic Vibrations; Sudden Changes of Amplitude.

Rigorously quasi-elastic forces, which are exactly proportional to the amplitude, even when it is very large, do not occur in nature. The actual binding forces are to be represented by power series in powers of the displacement. In most cases, a single additional term, along with the first power, is sufficient. Now there are two essentially different kinds of possible laws of force. If the force is *symmetric*, i.e. of equal magnitude at corresponding points on both sides of the position of equilibrium, or rest-point, then only *odd* powers may occur in the law of force; otherwise we have an *unsymmetric* law of force, and hence an *unsymmetric* vibration. We restrict ourselves here to the more common case of symmetric vibration, and add a term  $+\epsilon x^3$  to the quasi-elastic force  $-kx$ . We further assume that  $\epsilon$  is positive, which means that the binding becomes looser with increasing amplitude. The equation of motion for free vibrations of this system may be integrated readily; however, the result is expressible only by means of elliptic functions. The important fact is that the natural frequency *decreases* with increasing amplitude. This becomes evident qualitatively if we consider the actual restoring force replaced by a mean quasi-elastic force. On account of the form of the law of force, the restoring force will certainly become smaller with increasing amplitude  $x$ , and so, by equation (26) (p. 94), the natural frequency will also become less. The behaviour of such a system under the influence of a periodic external force is very peculiar. Here, even with no damping, the amplitude cannot increase without limit, for if resonance exists for infinitely small amplitudes, an increase in amplitude must cause a departure from the natural frequency, so that resonance no longer exists for larger amplitudes. The amplitude in the stationary state may be calculated by the following approximate method. The differential equation is



$$m \frac{d^2x}{dt^2} + kx - \epsilon x^3 = F_0 \cos \omega t. \quad . \quad . \quad . \quad (36)$$

The stationary solution is, at any rate, a periodic function of the time having a fundamental frequency  $\omega/2\pi$ , and is thus representable by a Fourier series in multiples of  $\omega$ . As a first approximation, we begin with the fundamental frequency and set

$$x_1 = a \cos \omega t.$$

The amplitude  $a$  is still to be determined. If we substitute this expression for  $x$  in equation (36), and make use of the trigonometrical formula

$$\cos^3 \omega t = \frac{1}{4} \cos 3\omega t + \frac{3}{4} \cos \omega t,$$

we obtain the equation

$$(-m\omega^2 + ka - \frac{3}{4}\epsilon a^3) \cos \omega t - \frac{\epsilon}{4} a^3 \cos 3\omega t = F_0 \cos \omega t.$$

If the fundamental vibration is to satisfy equation (36), we must have

$$\frac{3}{4} \frac{\epsilon}{m} a^3 + (\omega^2 - \omega_0^2) a + \frac{F_0}{m} = 0. \quad . \quad . \quad . \quad (37)$$

where we have put  $\omega_0$  for  $\sqrt{k/m}$ , the natural frequency at infinitesimal amplitude. The determination of the roots of this equation will be discussed below. At this point we wish to show how the next approximations are obtained. If equation (36) is solved for  $d^2x/dt^2$ , and if the first approximation is substituted in the right member, there results

$$\frac{d^2x}{dt^2} = \frac{F_0}{m} \cos \omega t - \frac{ka}{m} \cos \omega t + \frac{3}{4} \frac{\epsilon a^3}{m} \cos \omega t + \frac{\epsilon a^3}{4m} \cos 3\omega t.$$

By (37), this is equal to

$$-\omega^2 a \cos \omega t + \frac{\epsilon a^3}{4m} \cos 3\omega t.$$

Integration gives

$$\begin{aligned} x_2 &= a \cos \omega t - \epsilon \frac{a^3}{36m\omega^2} \cos 3\omega t \\ &= a \cos \omega t - \frac{\epsilon a^3 \omega_0^2}{36k\omega^2} \cos 3\omega t. \quad . \quad . \quad . \quad (38) \end{aligned}$$

This second approximation may then be substituted in equation (36). In this manner, any number of terms of the Fourier series may be calculated. The convergence of this process requires a separate

investigation. It may be said here, however, that the series converges for  $\epsilon$  small.

It remains to discuss the solution of the cubic equation for  $a$ . We

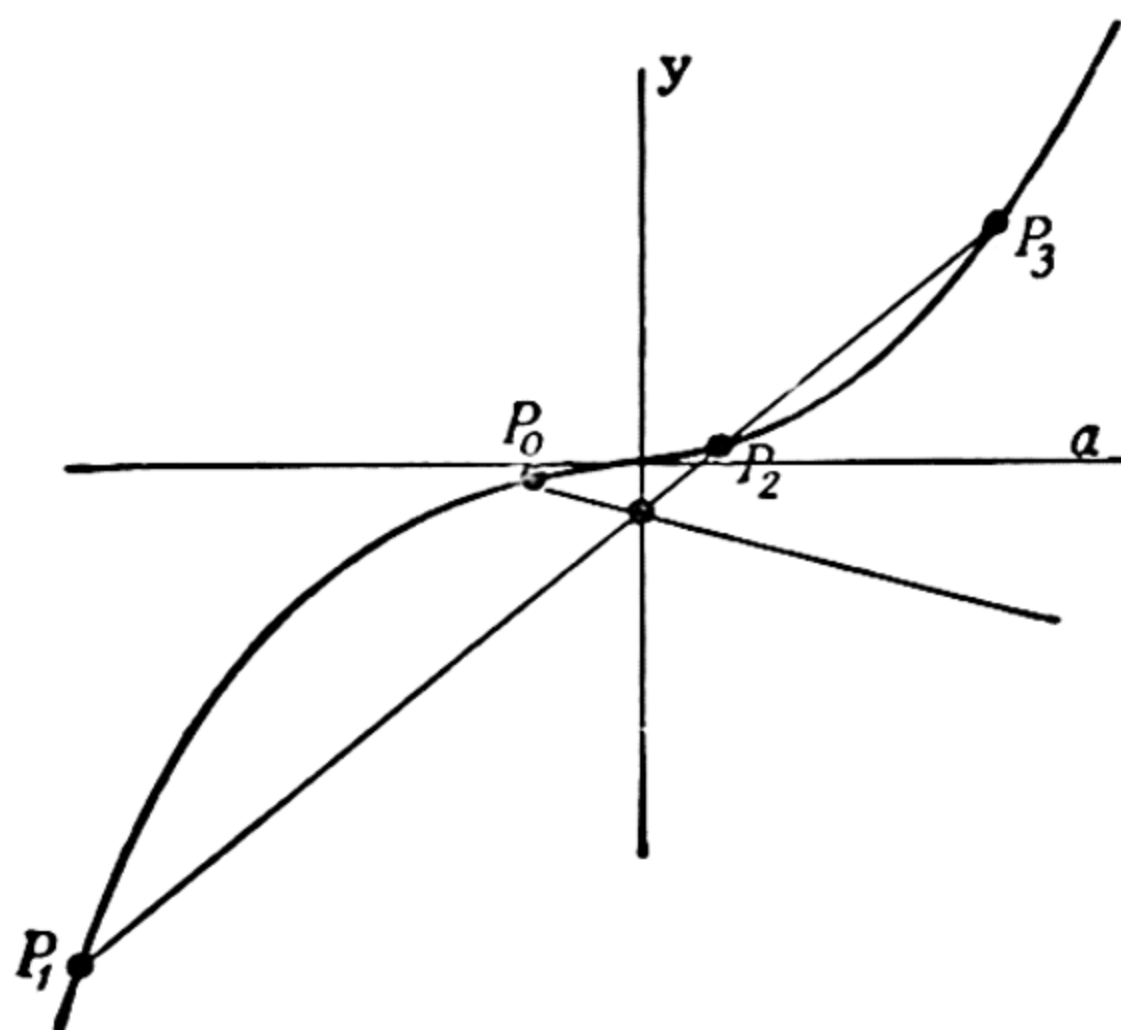


Fig. 8

obtain (fig. 8)  $a$  graphically in a  $y, a$  co-ordinate system as abscissa of the intersection of the cubical parabola

$$y = \frac{3\epsilon a^3}{4m\omega_0^2}$$

with a straight line

$$y = \left(1 - \frac{\omega^2}{\omega_0^2}\right)a - \frac{F_0}{\omega_0^2 m}.$$

If we start with  $\omega$  large, the slope of the straight line is negative, and there is only one point of intersection  $P_0$ . There is also but one intersection for  $\omega = \omega_0$ . If, now,  $\omega$  decreases, the straight line assumes a direction such that there are three points of intersection  $P_1, P_2, P_3$ , corresponding to three possible amplitudes. The resonance curve has the distorted form shown in fig. 9. A more thorough analysis shows that if we approach from the low-frequency side, the

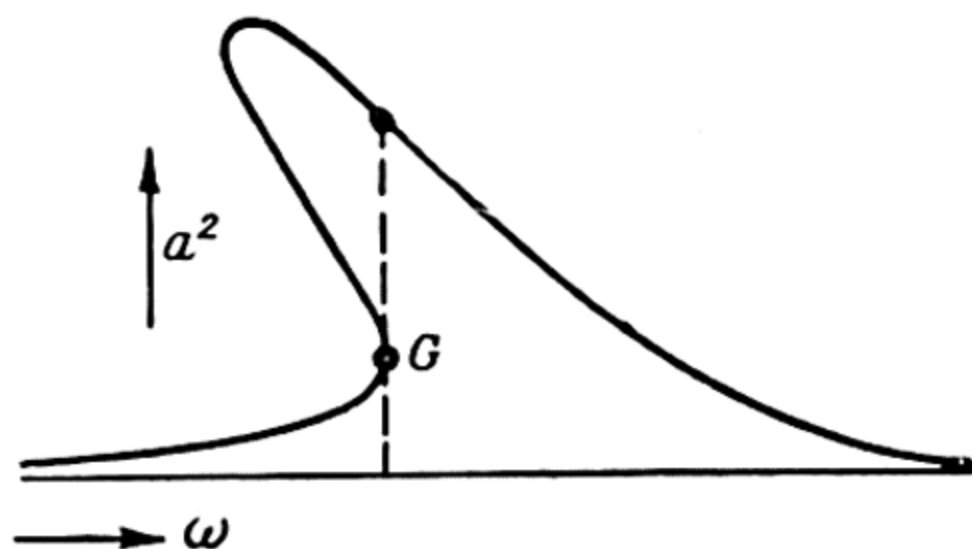


Fig. 9

amplitude corresponding to the lower branch is the stable one. As  $\omega$  increases we arrive at the limiting point  $G$ ; from this point onward, only the upper branch yields a real point of intersection in fig. 9. Thus, with continuous increase of  $\omega$ , the amplitude will jump suddenly from the lower branch to the upper, at or before  $G$ . Such discontinuities (jumps)

in amplitude, attributable to the non-harmonic nature of the restoring force, are frequently observed in vibration processes, both mechanical and electrical.

*Ex. 40.* Treat in similar manner the unsymmetric non-harmonic vibration having  $F' = -kx + \epsilon x^2$  as the law of force, and determine the mean value of the displacement.

## 11. Mechanics of a Constrained Particle. The Simple Pendulum.

The motion of a particle is often limited by geometric conditions. If, for example, we consider a small mass hanging by a thread of length  $l$ , its motion is restrained to the extent that it cannot move beyond the surface of a sphere of radius  $l$ . The whole interior of the sphere, however, is accessible in this case. The geometric condition would be represented by the inequality  $|\mathbf{r}| \leq l$ . But the commonest case is one in which the motion is confined to a surface or to a curve (i.e. the intersection of two surfaces). The condition is thus given by one or by two equations. In the last analysis, however, the physical attainment of these geometric conditions is realized by means of "forces". In the given example, these forces are the molecular forces between the particles of the thread and those at the junction of thread and weight. However, we need not concern ourselves with the atomic interpretation of these constraining forces. It is sufficient to replace the geometric condition by an equivalent constraining force in order to reduce the mechanics of a constrained particle to the case of the free particle.

An essential characteristic of all constraining forces is that they are perpendicular to the surface to which the motion is confined by these forces. If this were not the case, the constraints alone (i.e. the mere existence of a surface upon which the particle rests) could set it in motion without the action of externally impressed forces. This is contrary to everyday experience. For example, this would mean that a sphere resting on a smooth horizontal plane could set itself in motion. It is true that a tangential force may exist if the surface is rough, i.e. if there is friction between the particle and the surface; but this frictional force is inherently not a constraining force. If the magnitude of the constraining force is  $Z$ , and if  $\mathbf{n}$  is the unit normal vector to the surface, the equations of motion become

$$m \frac{d^2 \mathbf{r}}{dt^2} = \mathbf{F} + Z\mathbf{n}. \quad . \quad . \quad . \quad . \quad . \quad (39)$$

If the equation of the surface is  $G(x, y, z) = 0$ , then the gradient of the function  $G$  also has the direction of the normal. We may therefore write the constraining force as

$$\mathbf{Z} = \lambda \text{ grad } G, \quad . \quad . \quad . \quad . \quad . \quad (40)$$



where  $\lambda$  is a scalar quantity, as yet undetermined, and depending upon the co-ordinates. Thus the equation of motion is

$$m \frac{d^2 \mathbf{r}}{dt^2} = \mathbf{F} + \lambda \text{grad } G, \quad . . . . . (41)$$

or, in components,

$$m\ddot{x} = F_x + \lambda \frac{\partial G}{\partial x}, \quad m\ddot{y} = F_y + \lambda \frac{\partial G}{\partial y}, \quad m\ddot{z} = F_z + \lambda \frac{\partial G}{\partial z}. \quad (41')$$

The quantity  $\lambda$  has entered as a new unknown. But since the equation of the surface  $G = 0$  must also be satisfied, there are as many equations as there are unknowns. The same is true for the case of a particle constrained to move on a curve which is the intersection of two surfaces  $G_1 = 0$  and  $G_2 = 0$ . The system of equations then consists of the equation of motion

$$m\ddot{\mathbf{r}} = \mathbf{F} + \lambda_1 \text{grad } G_1 + \lambda_2 \text{grad } G_2, \quad . . . (41'')$$

containing the two unknown quantities  $\lambda_1$  and  $\lambda_2$ , along with the equations of the two surfaces.

Since the motion is confined to the surface at all times, every element of path  $d\mathbf{r}$  is perpendicular to the constraining force. For this reason no work is done by the constraints, and they do not contribute to the energy balance. In particular, if we are dealing with conservative forces, the equation of energy has the form derived for the case of the free particle:

$$\frac{1}{2}mv^2 + U = C. \quad . . . . . (42)$$

If a given problem involves but two co-ordinates—the case of motion in a plane—which are further connected by the equation of a curve to which the motion of the particle is confined, then the energy principle itself is sufficient for the integration; for one co-ordinate may be expressed in terms of the other by using the equation of the curve and the resulting derivatives, giving a differential equation in a single variable.

An instance of this kind is furnished by the simple, or plane mathematical pendulum. We understand this to mean a particle  $m$  fastened to one end of a weightless thread of length  $l$ . The thread restricts the motion of the particle to the inside of a sphere. We confine the discussion to plane motion, which may be realized, for example, by drawing the pendulum aside a given distance and then releasing it from rest. The plane of the motion is then determined by the initial position and equilibrium position (vertical) of the thread. The path still remains plane if the particle is given an initial velocity along the tangent to the circle cut from the sphere by the above plane.

On the other hand, the orbit is a space curve if the particle receives

an impulse the line of which does not lie in the plane. The conservation of energy alone is not sufficient for the solution of the motion of a space pendulum; an additional first integral of the equations of motion is needed. In this instance, the generalized Law of Areas (cf. Ex. 36, p. 89) will serve. This case will not be treated further here.

Returning to the discussion of the plane pendulum, the conservation of energy may be stated—using polar co-ordinates, in which the constraint is expressed simply by  $r = l$ ,

$$\frac{m}{2} l^2 \dot{\phi}^2 + mgl(1 - \cos \phi) = C, \quad . . . . (43)$$

the potential of gravity being given by  $mgz$ . As long as we rule out the case where the pendulum turns completely over, there are always two points at which the motion reverses, and where  $\dot{\phi}$  becomes zero. If we call the maximum amplitude, corresponding to such a reversal point, the angle  $\alpha$ , then, since  $\dot{\phi} = 0$ , (43) gives for this point

$$mgl(1 - \cos \alpha) = C,$$

and the energy equation becomes

$$\dot{\phi}^2 + \frac{2g}{l}(\cos \alpha - \cos \phi) = 0. \quad . . . . (44)$$

The integration is not difficult for small amplitudes. If we develop the cosines in series, and drop all terms beyond the squares, we obtain

$$\frac{d\phi}{dt} = \sqrt{\frac{g}{l}}(a^2 - \phi^2),$$

from which

$$\frac{d\phi}{\sqrt{a^2 - \phi^2}} = \sqrt{\frac{g}{l}} dt. \quad . . . . (45)$$

Integration gives

$$\sin^{-1}\left(\frac{\phi}{a}\right) = \sqrt{\frac{g}{l}} t + k \quad \text{or} \quad \phi = a \sin\left(\sqrt{\frac{g}{l}} t + k\right). \quad . (46)$$

The constant of integration  $k$  may be made zero by specifying that  $t = 0$  when  $\phi = 0$ , i.e. when the particle passes through the position of rest. We thus have a simple harmonic vibration of frequency

$$\nu = \frac{1}{2\pi} \sqrt{\frac{g}{l}}. \quad . . . . (47)$$

The period is thus

$$T = 2\pi \sqrt{\frac{l}{g}}. \quad . . . . (47')$$

The actual equation of motion may be obtained by differentiating equation (43) with respect to  $t$ ; this equation might also have been derived directly. It is

$$\ddot{\phi} + \frac{g}{l} \sin \phi = 0. \quad . \quad . \quad . \quad . \quad . \quad (48)$$

The first approximation putting  $\sin \phi = \phi$  gives the differential equation of simple harmonic motion. If the sine is developed to one additional term, we obtain an equation exactly like that of non-harmonic vibration, which was treated at length in § 10 (p. 100),\*

$$\ddot{\phi} + \frac{g}{l} \phi - \frac{g}{l} \frac{\phi^3}{6} = 0. \quad . \quad . \quad . \quad . \quad . \quad (49)$$

The pendulum, then, represents, for moderate amplitudes, a simple case of non-harmonic vibration. The rigorous integration for arbitrarily large amplitudes leads to elliptic functions. Thus, if we start with the energy equation (44) (p. 105) and replace  $\cos \phi$  by  $1 - 2 \sin^2(\phi/2)$  and  $\cos \alpha$  by  $1 - 2 \sin^2(\alpha/2)$ , we obtain, after separation of the variables,

$$\frac{d\phi}{\sqrt{\sin^2 \frac{\alpha}{2} - \sin^2 \frac{\phi}{2}}} = 2\sqrt{\frac{g}{l}} dt. \quad . \quad . \quad . \quad . \quad (50)$$

This is the exact equation corresponding to (45) (p. 105).

The integral is elliptic, and cannot be expressed in terms of elementary functions; it is, however, readily brought into the normal form of an elliptic integral of the first kind. If we put

$$\sin \frac{\phi}{2} = \sin \frac{\alpha}{2} \sin u, \quad . \quad . \quad . \quad . \quad . \quad (51)$$

we obtain

$$\frac{du}{\sqrt{1 - \sin^2 \frac{\alpha}{2} \sin^2 u}} = \sqrt{\frac{g}{l}} dt.$$

Integrating between  $t = 0$  and  $t = t$ , with the initial condition that for  $t = 0$ ,  $\phi = 0$  and therefore  $u = 0$ ,

$$\int_0^u \frac{du}{\sqrt{1 - \sin^2 \frac{\alpha}{2} \sin^2 u}} = F\left(\frac{\alpha}{2}, u\right) = \sqrt{\frac{g}{l}} t. \quad . \quad . \quad . \quad (52)$$

This integral is evaluated in many tables, where it is given in terms of  $u$  and  $\alpha/2$ . See, for example, E. Jahnke and F. Emde, *Tables of Functions* (New York, Dover Publications, 1943). The graphs on p. 61 of this book show vividly the departures from the limiting case of  $\alpha = 0$ .

\* There we also took into account the presence of a periodic external driving force.



The inverse of this integral is the *Jacobian Function*

$$\text{am} \left( \sqrt{\frac{g}{l}} t, \sin \frac{\alpha}{2} \right),$$

or 
$$u = \text{am} \left( \sqrt{\frac{g}{l}} t, \sin \frac{\alpha}{2} \right),$$

and 
$$\sin \frac{\varphi}{2} = \sin \frac{\alpha}{2} \text{sn} \left( \sqrt{\frac{g}{l}} t, \sin \frac{\alpha}{2} \right). \quad . . . . . (53)$$

The function sn (sine amplitude) is given in tables, with the help of which the course of the solution of (53) may be traced.

We inquire, finally, the magnitude of the restraining force which holds the particle in its circular path. According to Newton's Third Law concerning the equality of action and reaction, this force is equal to that with which the particle pulls on the cord. Since it is normal to the given surface, it must be radial. In the neighbourhood of the equilibrium point, the constraining force is certainly directed towards the centre.

The magnitude is readily found by writing the equation of motion obtained by resolving along the normal. Taking, as usual, the direction toward the centre of curvature as the positive direction of normal acceleration, we have

$$m \frac{v^2}{\rho} = m \frac{l^2 \dot{\varphi}^2}{l} = -mg \cos \varphi + Z.$$

Putting for  $\dot{\varphi}^2$  its value from the energy equation (44) (p. 105), we obtain

$$Z = mg(3 \cos \varphi - 2 \cos \alpha). \quad . . . . . (54)$$

The constraining force  $Z$  changes its sign when

$$3 \cos \varphi_m = 2 \cos \alpha.$$

This equation has a root which is compatible with the condition that  $\varphi < \alpha$  only when  $\alpha > \pi/2$ . This means that as long as  $\alpha$  does not exceed  $\pi/2$ , the thread remains taut. Arbitrarily large amplitudes are, however, not possible with a pendulum consisting of a particle suspended by a thread, for there is a point (viz. where  $\varphi = \varphi_m$ ) where the pendulum collapses. If we take, for example, an amplitude  $\alpha = \pi$  (this means that the particle just reaches the top of its circle) we have

$$\cos \varphi_m = -\frac{2}{3}, \quad \varphi_m = 131^\circ 49'.$$

The reason for this peculiarity is that a thread does not secure the geometric condition  $r = l$  completely; it excludes points for which  $r > l$ , but not points for which  $r < l$ . A satisfactory mechanism is provided by replacing the thread by a rigid, weightless rod.

*Ex. 41.* For what curve is the period of plane pendulum motion independent of the amplitude? [Introduce the length of arc  $s$  as a function of  $z = l(1 - \cos \varphi)$ ].

*Ex. 42.* A motor-cyclist is able to take a curve of 10 m. radius on a flat, unbanked road at 20 km./hr. without skidding. With what speed must he ride on the vertical inside surface of a circular cylinder of 5 m. radius, made of the same material, in order that he be able to ride in a horizontal path without slipping down? Recall that the traction force due to friction is proportional to the normal pressure and independent of the speed.

## CHAPTER VI

### GENERAL THEOREMS ON THE MECHANICS OF SYSTEMS OF PARTICLES

We represent a physical system schematically by means of a number of particles and apply the results of Chapter V to the individual particles. From these results we can deduce laws referring to the behaviour of the system as a whole. The range of validity of these laws is almost unlimited, since we can, in the last analysis, consider any body to be composed of a system of enormously many particles—electrons and protons. The forces acting upon a single particle are of two kinds: either they originate in the other particles of the system—*internal* forces—or they have their origin outside the system—*external* forces. *Newton's Third Law of the equality of action and reaction* is of especial importance here. We understand this law to have the following meaning: If a particle  $P_1$  exerts on  $P_2$  an attractive force directed toward  $P_1$ , then it is at once evident that  $P_2$  attracts  $P_1$  toward  $P_2$  with the same force, i.e. the force vectors of the two internal forces resulting from the mutual attraction are equal and opposite.

#### 1. Theorem concerning the Motion of the Centre of Mass.

We start with a system of  $N$  particles. The  $k$ th particle is subject to the following forces: A number of external forces which we replace by their resultant  $\mathbf{F}_k$ ; further, the force  $\mathbf{F}_{1k}$  due to the presence of the first particle,  $\mathbf{F}_{2k}$  from the second, &c.—in general,  $\mathbf{F}_{ik}$  from the  $i$ th particle. The equation of motion for the  $k$ th particle is thus

$$m_k \frac{d^2 \mathbf{r}_k}{dt^2} = \mathbf{F}_k + \sum_i \mathbf{F}_{ik}. \quad . \quad . \quad . \quad . \quad (1)$$

There are  $N$  such equations, one for each particle. Imagine them all written down and added together:

$$\sum_k m_k \frac{d^2 \mathbf{r}_k}{dt^2} = \sum_k \mathbf{F}_k + \sum_k \sum_i \mathbf{F}_{ik}. \quad . \quad . \quad . \quad . \quad (2)$$

Since internal forces having both subscripts alike do not exist, according to our notation, the combinations  $k = i$  are to be excluded from the double sum. Now for every force  $\mathbf{F}_{jk}$ —the force exerted by the

$j$ th particle on the  $k$ th—there corresponds a force  $\mathbf{F}_k$ ,—that exerted by the  $k$ th particle on the  $j$ th—and these two forces are equal and opposite. Hence the double sum vanishes, and *the internal forces of the system cancel out in the summation*. There remains in the right member only the vector sum of the external forces acting on the individual particles. We now define the *centre of mass* of a system to be a point whose radius vector  $\bar{\mathbf{r}}$  (referred to an arbitrary centre) multiplied by the total mass of the system is equal to the vector sum of the products of individual radius vectors of the separate particles with the corresponding masses:

$$M\bar{r} = \sum_k m_k r_k. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

If we substitute this expression in equation (2), we have the theorem

$$M \frac{d^2 \bar{\mathbf{r}}}{dt^2} = \sum_k \mathbf{F}_k. \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

*The centre of mass of a system moves as if the entire mass of the system were concentrated there, with the resultant of the externally applied forces acting at that point.* In particular, if there are no external forces, the centre of mass remains at rest or in a state of uniform rectilinear motion. As is well known, this theorem is the basis of the explanation of recoil phenomena. For example, if a shot is fired from a cannon standing upon a smooth horizontal plane, then the gun must spring back with a velocity such that the common centre of mass of cannon and projectile remains in the vertical line through the point of firing; for, neglecting friction of the gun with the ground, the only external force is gravity, which has no horizontal component.

Since the most universal external force is that of gravity, the centre of mass is commonly referred to as the *centre of gravity*. Another name for it in equally general use is the *centre of inertia*. The following elementary considerations are useful in determining this point: If  $\bar{r}$  is the radius vector of the centre of gravity of two particles  $m_1$  and  $m_2$ , then

$$(m_1 + m_2) \bar{r} = m_1 r_1 + m_2 r_2,$$

**OR**

$$m_1(\bar{r} - r_1) = m_2(r_2 - \bar{r});$$

this means that the vectors  $\bar{r} - r_1$  and  $r_2 - \bar{r}$  are parallel. But since they have the terminus of  $\bar{r}$  in common, the three points  $m_1$ ,  $m_2$  and the centre of gravity are collinear. The position of the centre of gravity is determined by

$$\frac{|\bar{r} - r_1|}{|r_2 - \bar{r}|} = \frac{m_2}{m_1}.$$

We thus have the rule: The centre of gravity of two particles  $m_1$  and  $m_2$  divides the distance between the particles in the ratio of the two masses, the centre of gravity being nearer the larger mass. If, now, a third particle be added to the system, the centre of gravity of the set will be the centre of gravity of  $m_3$  and the original centre of gravity, where both  $m_1$  and  $m_2$  may be considered concentrated. It is readily seen that the centre of gravity, found in this way, is independent of



the order in which the particles are taken. The procedure is similar for additional particles.

*Ex. 43.* Two particles  $m_1$  and  $m_2$  are in equilibrium when at a distance  $a$  apart. If the distance be altered a small amount  $x$ , there arise restoring forces proportional to  $x$ . Calculate the characteristic frequencies of the system.

## 2. Angular Momentum of a System of Particles.

According to § 5, p. 88, the angular momentum (or moment of momentum)  $\phi = m[\mathbf{r}\dot{\mathbf{r}}]$  of a particle is equal to  $2m$  times the "areal" velocity of the radius vector. The vector sum  $\sum m_k[\mathbf{r}_k\dot{\mathbf{r}}_k]$ , called the *total angular momentum* (or *moment of momentum*) of a system of particles, is one of the most important quantities in mechanics. We shall now investigate the properties of this quantity. For this purpose we multiply the equations of motion for a particle vectorially by  $\mathbf{r}_k$  and sum over all the particles. The result is

$$\sum_k m_k \left[ \mathbf{r}_k \frac{d^2 \mathbf{r}_k}{dt^2} \right] = \sum_k [\mathbf{r}_k \mathbf{F}_k] + \sum_k \sum_j [\mathbf{r}_k \mathbf{F}_{jk}]. \quad (5)$$

As shown at p. 89, the left member represents the time derivative of the quantity  $\sum m_k[\mathbf{r}_k\dot{\mathbf{r}}_k]$ , that is, of the total angular momentum. Further, the vector product of the radius vector to the point of application of a force by the force vector is called the *moment of the force*  $\mathbf{F}_k$ . We denote it by  $\mathbf{M}_k$ . The magnitude of  $\mathbf{M}_k$  corresponds to the product of force by lever arm, in relation to turning about  $O$ . The total moment of the external forces is given by  $\sum [\mathbf{r}_k \mathbf{F}_k]$ . The second term in the right member of (5), which represents the resultant of the moments of the internal forces, vanishes if the internal forces between two particles have the direction of the line joining the particles, i.e. if the forces are central. Thus, since  $\mathbf{F}_{jk} = -\mathbf{F}_{kj}$ , we have for any pair of particles

$$[\mathbf{r}_k \mathbf{F}_{jk}] + [\mathbf{r}_j \mathbf{F}_{kj}] = [(\mathbf{r}_k - \mathbf{r}_j) \mathbf{F}_{jk}]. \quad (6)$$

But the vector product on the right vanishes, since we are assuming that  $\mathbf{F}_{jk}$  is in the direction of  $\mathbf{r}_k - \mathbf{r}_j$ . There remains, therefore,

$$\frac{d}{dt} \sum_k m_k [\mathbf{r}_k \dot{\mathbf{r}}_k] = \frac{d\mathbf{P}}{dt} = \sum_k [\mathbf{r}_k \mathbf{F}_k] = \mathbf{M}. \quad (7)$$

*For a system of particles in which the forces between any two particles are in the direction of the line joining these particles, the rate of change of the total angular momentum is equal to the sum of the moments of the applied forces.*

The limitation made above is actually of little importance. From considerations of symmetry, it is difficult to imagine a force acting between two points which does not coincide in direction with the line joining them, for there is no other pre-eminent direction. If the Law of *Biot-Savart* (p. 303) seems an excep-

tion, it must be remembered that this law deals with the force between a magnet-pole and an elementary *segment* (i.e. *not* a point or particle) of an electrical conductor.

In the particular case in which there are no external forces acting, or if the total moment of the forces vanishes, then, according to equation (7), the total angular momentum of the system remains constant. In this form the theorem explains a large variety of phenomena of everyday life, e.g. the method by which a child sets a swing in motion, the ability of a falling cat to right itself before landing, the familiar turntable experiments, &c. This law finds one of its most beautiful applications in explaining the *Einstein-de Haas* Effect in Magnetism (cf. p. 464).

In general, the value of the total angular momentum depends upon the choice of the reference point  $O$ . If, however, the centre of gravity of the system is at rest, this quantity becomes independent of the choice of  $O$ . If we denote the radius vector to a new centre  $O'$  by  $\mathbf{r}_0$  and a radius vector emanating from  $O'$  by  $\mathbf{r}_k'$ , then

$$\mathbf{r}_k = \mathbf{r}_0 + \mathbf{r}_k'.$$

The angular momentum, referred to  $O$ , is

$$\mathbf{P} = \sum_k m_k [\mathbf{r}_k \dot{\mathbf{r}}_k],$$

and referred to  $O'$ ,

$$\mathbf{P}' = \sum_k m_k [\mathbf{r}_k' \dot{\mathbf{r}}_k'].$$

If we put  $\mathbf{r}_0 + \mathbf{r}_k'$  for  $\mathbf{r}_k$  in  $\mathbf{P}$ , we have, on account of  $\dot{\mathbf{r}}_0 = 0$ ,

$$\mathbf{P} = \sum_k m_k [(\mathbf{r}_0 + \mathbf{r}_k') \dot{\mathbf{r}}_k'] = [\mathbf{r}_0 \cdot \sum_k m_k \dot{\mathbf{r}}_k'] + \mathbf{P}'.$$

But the first term vanishes if

$$\sum m_k \frac{d\mathbf{r}_k'}{dt} = 0,$$

i.e. if the centre of gravity is at rest.

### 3. Total Energy of a System of Particles.

Let us multiply the equation of motion of the  $k$ th particle scalarly with  $d\mathbf{r}/dt$ , and sum over all the particles. Then

$$\sum_k m_k \frac{d^2 \mathbf{r}_k}{dt^2} \frac{d\mathbf{r}_k}{dt} = \frac{d}{dt} \frac{1}{2} \sum_k m_k \left( \frac{d\mathbf{r}_k}{dt} \right)^2 = \sum_k \mathbf{F}_k \frac{d\mathbf{r}_k}{dt} + \sum_k \sum_{j \neq k} \mathbf{F}_{jk} \frac{d\mathbf{r}_k}{dt}. \quad (8)$$

Integrating between the times  $t_0$  and  $t$ :

$$\begin{aligned} \frac{1}{2} \sum_k m_k \left( \frac{d\mathbf{r}_k}{dt} \right)^2_t - \frac{1}{2} \sum_k m_k \left( \frac{d\mathbf{r}_k}{dt} \right)^2_{t_0} \\ = \int_{r_k(t_0)}^{r_k(t)} \sum_k \mathbf{F}_k d\mathbf{r}_k + \int_{r_k(t_0)}^{r_k(t)} \sum_k \sum_j \mathbf{F}_{jk} d\mathbf{r}_k. \quad \dots (9) \end{aligned}$$

The left member represents the total change in kinetic energy of the

system, the right member gives the work done by the internal and external forces. But it is by no means the case that the work done by the internal forces cancels out in calculating the energy, as one might expect it to do in view of the results of the two preceding sections.

The kinetic energy may be divided into two parts, each of which has a physical meaning. If we introduce a second co-ordinate system, whose origin  $O'$  is at the centre of gravity of the system, and if we denote all radius vectors referred to this system by primes, we have

$$\mathbf{r}_k = \bar{\mathbf{r}} + \mathbf{r}_k'.$$

Then, identically,

$$\begin{aligned} \Sigma_k \frac{1}{2} m_k \left( \frac{d\mathbf{r}_k}{dt} \right)^2 \\ = \frac{1}{2} \left( \frac{d\bar{\mathbf{r}}}{dt} \right)^2 \Sigma_k m_k + \frac{d\bar{\mathbf{r}}}{dt} \Sigma_k m_k \frac{d\mathbf{r}_k'}{dt} + \frac{1}{2} \Sigma_k m_k \left( \frac{d\mathbf{r}_k'}{dt} \right)^2. \end{aligned} \quad (10)$$

The second summand on the right vanishes, however, since  $\Sigma m_k \mathbf{r}_k / M$  is, by equation (3) (p. 109), the radius vector of the centre of gravity—and this, by hypothesis, is zero in the primed co-ordinates. The first term on the right represents the kinetic energy of the system, considering the entire mass to be concentrated at the centre of gravity. The last term gives the kinetic energy of motion of the system referred to the centre of gravity, when considered as at rest. Thus, we may say: *The total kinetic energy is equal to the translational kinetic energy of the entire mass, considered concentrated at the centre of gravity, plus the energy of motion of the parts of the system relative to the centre of gravity.*

We further assume that the internal forces are such that they are derivable from a potential. The potential of the force operating between the points  $j$  and  $k$  is a function of the distance between the two points, and therefore of their co-ordinates:

$$U_{jk} = U_{jk}(\mathbf{r}_{jk}) = U_{jk}(\sqrt{(x_j - x_k)^2 + (y_j - y_k)^2 + (z_j - z_k)^2}). \quad (11)$$

The force acting on  $k$  is obtained by taking  $j$  to be fixed, and considering  $k$  to move in the potential field given by the point function  $U_{jk}$ ; i.e. we consider the co-ordinates of  $j$  to be fixed, those of  $k$  to be variable. Then,

$$\mathbf{F}_{jk} = -i \frac{\partial U_{jk}}{\partial x_k} - j \frac{\partial U_{jk}}{\partial y_k} - k \frac{\partial U_{jk}}{\partial z_k} = -\text{grad}_k U_{jk}. \quad (12)$$

in like manner,

$$\mathbf{F}_{kj} = -i \frac{\partial U_{jk}}{\partial x_j} - j \frac{\partial U_{jk}}{\partial y_j} - k \frac{\partial U_{jk}}{\partial z_j} = -\text{grad}_j U_{jk} = -\mathbf{F}_{jk}. \quad (13)$$



The work done in causing small displacements of  $j$  and  $k$  is

$$\mathbf{F}_{jk} d\mathbf{r}_k + \mathbf{F}_{kj} d\mathbf{r}_j = - \left( \frac{\partial U_{jk}}{\partial x_k} dx_k + \frac{\partial U_{jk}}{\partial y_k} dy_k + \frac{\partial U_{jk}}{\partial z_k} dz_k + \frac{\partial U_{jk}}{\partial x_j} dx_j + \frac{\partial U_{jk}}{\partial y_j} dy_j + \frac{\partial U_{jk}}{\partial z_j} dz_j \right) = -dU_{jk}. \quad (14)$$

The negative of the sum of  $\mathbf{F}_{jk} d\mathbf{r}_k$  and  $\mathbf{F}_{kj} d\mathbf{r}_j$  is therefore obtained by forming the total differential of  $U_{jk}$ , defined as a function of the six co-ordinates of the two points, in (11). If, then, we wish to introduce the internal potential into the right member of equation (9) (p. 111), we must write

$$\sum_k \sum_j \mathbf{F}_{jk} d\mathbf{r}_k = -\frac{1}{2} \sum_k \sum_j dU_{jk}, \quad (k \neq j). \quad (15)$$

It is readily seen that the factor  $1/2$  enters: If we start with point 1, and calculate the mutual energy  $U_{jk}$  between this and all the other points,  $k$  runs from 2 to  $N$ ; but when we take point 2, we must start counting with 3, since the mutual effect of points 1 and 2 was already taken into account in dealing with point 1, and so on. Thus, in extending the summation over all combinations  $j$  and  $k$ , we must divide by two.

If the external forces have also a potential, the energy equation (9) becomes

$$T + \sum_k U_k + \frac{1}{2} \sum_k \sum_j U_{jk} = T^{(0)} + \sum_k U_k^{(0)} + \frac{1}{2} \sum_k \sum_j U_{jk}^{(0)} = \text{const.} \quad (j \neq k), \quad (16)$$

where  $T$  denotes the kinetic energy. *The sum of the kinetic energy and of the external and internal potential energy of a system is constant, if the external as well as the internal forces are conservative.*

**Ex. 44.** A solid sphere of radius  $r$  and density  $\rho$  rolls without slipping down a plane inclined at an angle  $\alpha$  to the horizontal. What is the velocity of its centre after rolling a distance  $s$ , and how does this velocity compare with that which would be attained by the sphere in sliding without friction the same distance?

**Ex. 45.** Give the theory of the central impact of two spheres, masses  $m_1$  and  $m_2$ , whose velocities are along the line joining their centres: (a) for the case of perfectly elastic spheres, where no mechanical energy is lost in the impact (i.e. converted into heat); (b) for a case of inelastic impact in which  $m_2$  acquires the internal energy  $\epsilon$ .

#### 4. The Principle of Virtual Displacements, d'Alembert's Principle and the Lagrangian Equations of the First Kind.

Let any number of forces be applied to the  $k$ th particle of a system. Assume that the individual particles are completely free, i.e. there are no constraints imposed by geometric conditions. Then the condition that the system is not to move under the influence of the forces is

evidently the condition for *equilibrium*, viz. that the resultant  $\mathbf{F}_k$  is to vanish:

$$\mathbf{F}_k = 0.$$

If we imagine each particle displaced from its rest position by a vector  $\delta\mathbf{r}_k$ , then  $\sum \mathbf{F}_k \delta\mathbf{r}_k$  represents the total work done in such a *virtual* displacement. Since each  $\mathbf{F}_k$  vanishes, this sum is evidently zero. *If a system of free particles is in equilibrium, the total work done by all forces, internal and external, in a small virtual displacement is zero.* If the forces have a potential  $U = \sum U_k$ , then

$$\sum \mathbf{F}_k \delta\mathbf{r}_k = -\sum \delta U_k = -\delta U = 0. \quad . . . (17)$$

The vanishing of the variation  $\delta U$  means, however, that the function has a stationary value. *In a state of equilibrium the potential energy of a system of free particles has a stationary value.*

At first sight these considerations seem trivial; their utility first appears when limitations are imposed in the form of geometric conditions; e.g. the distance between two particles is to remain constant, or certain particles are to be confined to given surfaces, &c. The constraining forces, as was shown in Chap. V (p.104), can do no work. Consequently the principle of virtual displacements holds also for the equilibrium of a system with constraints, provided we understand that the  $\mathbf{F}_k$  are the actual given forces. On the other hand, it is no longer true that  $\mathbf{F}_k$  must be zero; on the contrary, an applied force may be balanced by a constraining force. For example, the force of gravity acting on a mass which rests upon a horizontal supporting surface is balanced by the reaction of the surface. This difference between a free and a constrained system is expressed mathematically by the fact that the displacements  $\delta\mathbf{r}_k$  are no longer independent, but are such that certain auxiliary conditions are satisfied. If we write equation (17) in terms of the co-ordinates, which we designate by consecutive numbers instead of by  $x, y, z$ , then the sum

$$\sum_i F_i \delta x_i = 0 \quad . . . . . (17')$$

has  $3N$  terms, for a system of  $N$  particles. Each equation of condition reduces the number of  $3N$  arbitrary variables by one. For example, if one equation of condition exists, we can express the last displacement in terms of the others, and substitute in equation (17), which will then contain only independent displacements. However, the following procedure, devised by Lagrange, affords a much better general view of the process: let the conditions be given in the form of finite equations connecting the co-ordinates  $x_i$  (*holonomous* systems), which are independent of the time (*scleronomous* systems).

Let there be  $l$  equations of condition. The  $k$ th of these,

$$f_k(x_1, x_2, \dots x_{3N}) = 0,$$

gives upon differentiating,

$$\frac{\partial f_k}{\partial x_1} \delta x_1 + \frac{\partial f_k}{\partial x_2} \delta x_2 + \dots + \frac{\partial f_k}{\partial x_{3N}} \delta x_{3N} = 0. \quad \cdot \cdot \quad (18)$$

If we multiply these equations by certain provisionally undetermined multipliers  $\lambda_k$ , and add them to equation (17'), we have

$$\Sigma_i \left( F_i + \lambda_1 \frac{\partial f_1}{\partial x_i} + \lambda_2 \frac{\partial f_2}{\partial x_i} + \dots \lambda_l \frac{\partial f_l}{\partial x_i} \right) \delta x_i = 0. \quad (19)$$

In this sum of  $3N$  terms we can select the multipliers in such a way that the last  $l$  terms vanish. The remaining  $(3N - l)$  terms then contain only  $(3N - l)$  displacements, i.e. as many as there are independent quantities  $\delta x_i$  present. However, since these  $\delta x_i$  are independent of each other, their coefficients must vanish separately, in order that the sum be zero. For we may select the purely arbitrary displacements in such a way that, for example, all the  $\delta x_i$  except  $\delta x_n$  are zero, in which case the only remaining member of the sum is

$$\left( F_n + \lambda_1 \frac{\partial f_1}{\partial x_n} + \dots \lambda_i \frac{\partial f_i}{\partial x_n} \right) \delta x_n.$$

But since, by hypothesis,  $\delta x_n$  is to be different from zero, the expression in brackets must vanish. We thus obtain the system of equations

$$\left. \begin{aligned} F_1 + \lambda_1 \frac{\partial f_1}{\partial x_1} + \dots \lambda_l \frac{\partial f_l}{\partial x_1} &= 0 \\ F_2 + \lambda_1 \frac{\partial f_1}{\partial x_2} + \dots \lambda_l \frac{\partial f_l}{\partial x_2} &= 0 \\ . &. . . . . \\ F_{3N} + \lambda_1 \frac{\partial f_1}{\partial x_{3N}} + \dots \lambda_l \frac{\partial f_l}{\partial x_{3N}} &= 0 \end{aligned} \right\}. \quad . \quad . \quad . \quad (20)$$

Despite the fact that these equations are all of the same form, they have different origins.  $l$  of these equations hold because we purposely determine the  $l$  quantities  $\lambda_k$  in such way as to make them true; the remaining  $(3N - l)$  equations are true, however, because the determination of the  $\lambda_k$  in this way makes the  $(3N - l)$  displacements independent of one another. Besides these  $3N$  equations, we have also the  $l$  equations of condition, so that the number of equations is equal to  $(3N + l)$ , which is the same as the number of unknowns, viz. the  $3N$  co-ordinates  $x_i$  of the particles, and the  $l$  multipliers  $\lambda_i$ . On the other hand, if we take the constraining forces  $\mathbf{Z}_k$  which embody the geometric conditions, instead of the conditions themselves, we then have a free system for which, from above,

$$F_t + Z_t = 0. \quad . \quad . \quad . \quad . \quad . \quad (21)$$



Comparison with equation (20) shows that the constraining forces are represented by

$$Z_i = \lambda_1 \frac{\partial f_1}{\partial x_i} + \lambda_2 \frac{\partial f_2}{\partial x_i} + \dots + \lambda_l \frac{\partial f_l}{\partial x_i}. \quad (22)$$

Thus far we have been occupied only with *statics*, or equilibrium, but dynamics can be treated in a quite similar way. Denoting the resultant of the forces applied to the  $k$ th particle by  $\mathbf{F}_k$ , that of the constraint forces by  $\mathbf{Z}_k$ , we have

$$\mathbf{F}_k + \mathbf{Z}_k - m_k \frac{d^2 \mathbf{r}_k}{dt^2} = 0.$$

If a virtual displacement is again performed, then the constraints by themselves can do no work, and we must have  $\sum_k \mathbf{Z}_k \delta \mathbf{r}_k = 0$ . We thus obtain the relation known as *d'Alembert's Principle*:

$$\sum_k \left( \mathbf{F}_k - m_k \frac{d^2 \mathbf{r}_k}{dt^2} \right) \delta \mathbf{r}_k = 0. \quad (23)$$

Since the equations of condition limit the freedom of the  $\delta \mathbf{r}_k$ , we cannot immediately conclude from (23) that the individual terms of the sum vanish. We must, in fact, proceed exactly as in statics. Again we label the co-ordinates consecutively, remembering that each mass occurs three times, i.e.  $m_{3n} = m_{3n-1} = m_{3n-2}$ . The relation (23) then becomes

$$\sum (F_i - m_i \ddot{x}_i) \delta x_i = 0. \quad (24)$$

Let the  $l$  equations of condition again be finite equations. Differentiation gives

$$\frac{\partial f_k}{\partial x_1} \delta x_1 + \frac{\partial f_k}{\partial x_2} \delta x_2 + \dots + \frac{\partial f_k}{\partial x_{3N}} \delta x_{3N} = 0.$$

We again multiply each of these equations by a multiplier  $\lambda_k$  (for the present undetermined), and add them to equation (24), whence we have

$$\sum_i \left( F_i - m_i \ddot{x}_i + \lambda_1 \frac{\partial f_1}{\partial x_i} + \lambda_2 \frac{\partial f_2}{\partial x_i} + \dots + \lambda_l \frac{\partial f_l}{\partial x_i} \right) \delta x_i = 0. \quad (25)$$

Exactly as in the statical case, we can dispose of the  $l$  quantities  $\lambda_k$  so that the first  $l$  parentheses vanish. The  $(3N - l)$  remaining parentheses contain as many displacements as there are independent variables present. Since we may choose these  $\delta x_i$  at will, the remaining parentheses must also vanish separately. We then obtain a

system of equations known as *Lagrange's Equations of the First Kind*:

$$\left. \begin{aligned} m_1 \ddot{x}_1 &= F_1 + \lambda_1 \frac{\partial f_1}{\partial x_1} + \lambda_2 \frac{\partial f_2}{\partial x_1} + \dots + \lambda_l \frac{\partial f_l}{\partial x_1} \\ &\vdots \\ m_i \ddot{x}_i &= F_i + \lambda_1 \frac{\partial f_1}{\partial x_i} + \lambda_2 \frac{\partial f_2}{\partial x_i} + \dots + \lambda_l \frac{\partial f_l}{\partial x_i} \\ &\vdots \\ m_{3N} \ddot{x}_{3N} &= F_{3N} + \lambda_1 \frac{\partial f_1}{\partial x_{3N}} + \lambda_2 \frac{\partial f_2}{\partial x_{3N}} + \dots + \lambda_l \frac{\partial f_l}{\partial x_{3N}} \end{aligned} \right\} \quad (26)$$

In this case also we easily see that the terms added to  $F_i$  in the right member are the components of the constraining forces, which must be introduced as the forces corresponding to the geometric conditions, if we wish to treat the particles as free. The number of equations is the same as the number of unknowns. The Lagrangian equations contain, besides the  $3N$  co-ordinates of the particles, the  $l$  multipliers; on the other hand, we have, in addition to the  $3N$  Lagrange equations, the  $l$  equations of condition

$$F_k(x_1 \dots x_{3N}) = 0.$$

If the integration of the entire system of equations (26) (including the  $l$  equations of condition) can be performed, we obtain the components of the constraining forces, since the  $\lambda_k$  are also known. These forces of constraint are equal and opposite to the pressures or tensions exerted by the particles on the guiding mechanism which secures the geometric conditions. Such forces were studied for a single particle in the case of the pendulum. A pertinent example, where external forces are absent, is that of a particle whirled about at the end of a string tied to a fixed point in a horizontal plane, the force of gravity being balanced by the reaction of the plane. The *centripetal* force exerted by the string is the equilibrant of the *centrifugal* force of rotation of the particle.

Another way of looking at the question, convenient in many computations, is the following: The fundamental equation

$$\mathbf{F} - m \frac{d^2 \mathbf{r}}{dt^2} = 0$$

may be considered to express the equilibrium between the *applied force*  $\mathbf{F}$  and the *inertia force*  $-m(d^2 \mathbf{r}/dt^2)$ . In this instance, however, Newton's second law must be formulated to read:

*During the accelerated motion, a force  $-m(d^2 \mathbf{r}/dt^2)$  acts on the mass  $m$  and there is equilibrium between applied and inertial force. This view formally refers dynamics back to statics, and the considerations of*

statics may be applied to dynamics, so that we arrive again at (23). The difference in the two approaches may be illustrated by the example of a spring-type weighing machine, standing on which is a person who produces accelerated motion by flexing his knees. Without the introduction of inertial forces we have

$$m \frac{d^2z}{dt^2} = -mg + F,$$

where  $F$  is the restoring force in the spring. The other view says that the force  $F'$  exerted by the spring must equal the resultant of the applied forces. The latter are the gravitational force  $-mg$  and the inertial force  $-m(d^2z/dt^2)$ , so that

$$F' = -mg - m \frac{d^2z}{dt^2}.$$

According to the third law,  $F'$  must equal  $-F$ , so that both approaches lead to the same result. While the second formulation often yields the formal result more quickly, it must be remembered that the first one usually takes into account the physical situation more satisfactorily.

## 5. The Lagrangian Equations of the Second Kind for Arbitrary Co-ordinates (Generalized Co-ordinates).

The Lagrangian equations of the First Kind refer to rectangular co-ordinates. However, it is often more useful to specify the positions of the particles by means of other parameters. This almost invariably permits us to regard the equations of condition as equivalent to holding certain co-ordinates constant, and these co-ordinates then drop out of the conditional relationships. Although the introduction of conditions is also possible with any system of co-ordinates, we consider here the normal case where, in place of  $3N$  rectangular co-ordinates and  $l$  equations of condition, we have  $f = 3N - l$  independent parameters  $q_1, q_2, \dots, q_f$ . For example, let all points of the system be required to remain on the surface of a sphere. In this case the correct co-ordinates are spherical co-ordinates,  $r, \phi, \theta$ , and the variable quantities are only the angles  $\phi$  and  $\theta$ . It remains to express the equations of motion in these arbitrary co-ordinates. For this purpose we imagine the rectangular co-ordinates to be given as functions of the independent  $q_k$  (called the *generalized co-ordinates*)

$$x_i = x_i(q_1, q_2, \dots, q_f). \quad . \quad . \quad . \quad . \quad . \quad (27)$$

Further, we seek to replace the  $x_i, \delta x_i$  and  $\ddot{x}_i$  appearing in the d'Alembert equation

$$\Sigma_i (F_i - m_i \ddot{x}_i) \delta x_i = 0, \quad . \quad . \quad . \quad . \quad . \quad (24)$$



by the  $q_k$  and their time derivatives. Now

$$\delta x_i = \frac{\partial x_i}{\partial q_1} \delta q_1 + \frac{\partial x_i}{\partial q_2} \delta q_2 + \dots + \frac{\partial x_i}{\partial q_f} \delta q_f, \quad . \quad . \quad (28)$$

and similarly 
$$\dot{x}_i = \frac{\partial x_i}{\partial q_1} \dot{q}_1 + \frac{\partial x_i}{\partial q_2} \dot{q}_2 + \dots + \frac{\partial x_i}{\partial q_f} \dot{q}_f. \quad . \quad . \quad (28')$$

The expressions  $\ddot{x}_i(\partial x_i/\partial q_k)$  which result from substitution of (28) in (24), may be transformed as follows: it is verified at once that

$$\ddot{x}_i \frac{\partial x_i}{\partial q_k} = \frac{d}{dt} \left( \dot{x}_i \frac{\partial x_i}{\partial q_k} \right) - \dot{x}_i \frac{d}{dt} \left( \frac{\partial x_i}{\partial q_k} \right). \quad . \quad . \quad (29)$$

Now we can consider the  $\dot{x}_i$  to be functions of the  $2f$  variables  $q_k$  and  $\dot{q}_k$ .\*

Equation (28') then gives

$$\frac{\partial \dot{x}_i}{\partial \dot{q}_k} = \frac{\partial x_i}{\partial q_k}. \quad . \quad . \quad . \quad . \quad . \quad (30)$$

Then the first term in the right member of equation (29) becomes

$$\frac{d}{dt} \left( \dot{x}_i \frac{\partial x_i}{\partial q_k} \right) = \frac{d}{dt} \left( \dot{x}_i \frac{\partial \dot{x}_i}{\partial \dot{q}_k} \right) = \frac{d}{dt} \left( \frac{1}{2} \frac{\partial \dot{x}_i^2}{\partial \dot{q}_k} \right). \quad . \quad (31)$$

The second term may also be transformed in a way which leads to easily interpreted expressions. We have

$$\begin{aligned} \frac{d}{dt} \left( \frac{\partial x_i}{\partial q_k} \right) &= \frac{\partial^2 x_i}{\partial q_1 \partial q_k} \dot{q}_1 + \frac{\partial^2 x_i}{\partial q_2 \partial q_k} \dot{q}_2 + \dots + \frac{\partial^2 x_i}{\partial q_f \partial q_k} \dot{q}_f \\ &= \frac{\partial}{\partial q_k} \left( \frac{\partial x_i}{\partial q_1} \dot{q}_1 + \frac{\partial x_i}{\partial q_2} \dot{q}_2 + \dots \right) = \frac{\partial}{\partial q_k} \dot{x}_i, \quad . \quad (32) \end{aligned}$$

from which 
$$\dot{x}_i \frac{d}{dt} \left( \frac{\partial x_i}{\partial q_k} \right) = \frac{\partial}{\partial q_k} \left( \frac{1}{2} \dot{x}_i^2 \right). \quad . \quad . \quad . \quad (33)$$

Then, according to equations (28) and (29),

$$m_i \ddot{x}_i \delta x_i = \sum_k \left( \frac{d}{dt} \frac{\partial}{\partial \dot{q}_k} \frac{1}{2} m_i \dot{x}_i^2 - \frac{\partial}{\partial q_k} \frac{1}{2} m_i \dot{x}_i^2 \right) \delta q_k. \quad (33')$$

\* The introduction of the  $\dot{q}_k$  as new variables, in addition to the  $q_k$ , is apt to be misunderstood, but consideration of the following example of a similar kind may help to remove the difficulty. The only independent variable is the time, upon which the  $x_i$  and  $\dot{x}_i$  depend, through the medium of the  $q_k$ . However, if a function  $f(y, y')$  is to be differentiated with respect to the independent variable  $x$ , where  $y'$  stands for  $dy/dx$ , we use the formula

$$\frac{d}{dx} f(y, y') = \frac{\partial f}{\partial y} \frac{dy}{dx} + \frac{\partial f}{\partial y'} \frac{dy'}{dx}.$$

We further assume that the acting forces have potentials  $U_i$ , which we imagine to be given as functions of the  $q_k$ . Then

$$F_i \delta x_i = -\delta U_i = -\left(\frac{\partial U_i}{\partial q_1} \delta q_1 + \frac{\partial U_i}{\partial q_2} \delta q_2 + \dots + \frac{\partial U_i}{\partial q_r} \delta q_r\right). \quad (34)$$

If we now substitute expressions (33') and (34) in the statement of d'Alembert's principle, there results

$$\Sigma_k \left( \frac{d}{dt} \Sigma_i \frac{\partial}{\partial \dot{q}_k} \frac{1}{2} m_i \dot{x}_i^2 - \Sigma_i \frac{\partial}{\partial q_k} \frac{1}{2} m_i \dot{x}_i^2 + \Sigma_i \frac{\partial U_i}{\partial q_k} \right) \delta q_k = 0. \quad (35)$$

But  $\Sigma \frac{1}{2} m_i \dot{x}_i^2$  represents the total kinetic energy  $T$  of the system. Then, since the bracketed expressions must vanish individually, on account of the independence of the displacements  $\delta q_k$ , and since  $U$  depends only upon the position co-ordinates  $q_k$  and not upon  $\dot{q}_k$ , we obtain the following *Lagrangian Equations of the Second Kind* for conservative forces:

$$\frac{d}{dt} \frac{\partial (T - U)}{\partial \dot{q}_k} - \frac{\partial (T - U)}{\partial q_k} = 0. \quad \dots \dots (36)$$

The difference between the kinetic and potential energies, which is a function of the time (through the position co-ordinates and their time derivatives), is called the *Lagrangian Function*  $L$  of the system:

$$L = T - U. \quad \dots \dots (37)$$

Using this notation, equation (36) may be written

$$\frac{d}{dt} \frac{\partial L}{\partial \dot{q}_k} - \frac{\partial L}{\partial q_k} = 0. \quad \dots \dots (38)$$

If, in addition to the conservative forces  $F_i$ , non-conservative forces  $F'_i$  are also acting, then the work done by the latter in a virtual displacement is given by inserting the values of  $\delta x_i$  from equation (28) (p. 119):

$$\begin{aligned} \delta W &= \Sigma_i F'_i \delta x_i = \Sigma_i F'_i \left( \frac{\partial x_i}{\partial q_1} \delta q_1 + \frac{\partial x_i}{\partial q_2} \delta q_2 + \dots + \frac{\partial x_i}{\partial q_r} \delta q_r \right) \\ &= \left( \Sigma_i F'_i \frac{\partial x_i}{\partial q_1} \right) \delta q_1 + \left( \Sigma_i F'_i \frac{\partial x_i}{\partial q_2} \right) \delta q_2 + \dots + \left( \Sigma_i F'_i \frac{\partial x_i}{\partial q_r} \right) \delta q_r. \end{aligned} \quad (39)$$

The coefficients of the  $\delta q_k$  are called the components of the forces referred to the generalized co-ordinates  $q_k$ , and are denoted by  $F'_{qk}$ . With the admission of non-conservative forces, the Lagrangian equations of the second kind thus become

$$\frac{d}{dt} \frac{\partial L}{\partial \dot{q}_k} - \frac{\partial L}{\partial q_k} = F'_{qk}. \quad \dots \dots (40)$$

It is to be remembered that we have treated only the most common case, i.e. the holonomous-scleronomous system defined on p. 114. For a discussion of the cases in which the equations expressing the  $x_i$  as functions of the  $q_k$

(a) contain the time itself explicitly (holonomous-rheonomous system), and (b) are given in the form of non-integrable differential equations  $\delta x_i = a_1^{(i)} \delta q_1 + a_2^{(i)} \delta q_2 + \dots + a_f^{(i)} \delta q_f$  (non-holonomous system), consult the more complete textbooks on mechanics.

*Ex. 46.* Set up the equations of motion, in spherical co-ordinates, of a particle on the surface of the earth.

## 6. Generalized Momentum Co-ordinates. Hamilton's Equations.

In the case of holonomous-scleronomous systems, the  $x_i$  are given as functions of the  $q_k$  by finite equations which do not contain the time itself, and the kinetic energy  $T$  is a homogeneous quadratic function of the time derivatives of the  $q_k$ :

$$T = \sum_i \frac{1}{2} m_i \dot{x}_i^2 = \frac{1}{2} \sum m_i \left( \frac{\partial x_i}{\partial q_1} \dot{q}_1 + \frac{\partial x_i}{\partial q_2} \dot{q}_2 + \dots + \frac{\partial x_i}{\partial q_f} \dot{q}_f \right)^2$$

$$= \frac{1}{2} (P_{11} \dot{q}_1^2 + P_{22} \dot{q}_2^2 + \dots + 2P_{12} \dot{q}_1 \dot{q}_2 + \dots). \quad (41)$$

Now in rectangular co-ordinates, each component of the momentum is given by

$$p_{x_i} = m_i \dot{x}_i = \frac{d \frac{1}{2} m_i \dot{x}_i^2}{d \dot{x}_i} = \frac{\partial T}{\partial \dot{x}_i}. \quad (42)$$

In analogy with this, we call the following quantities the generalized momentum components:

$$p_k = \frac{\partial T(q_k, \dot{q}_k)}{\partial \dot{q}_k} = P_{1k} \dot{q}_1 + P_{2k} \dot{q}_2 + \dots + P_{fk} \dot{q}_f, \quad (43)$$

although these quantities may or may not have the dimensions of momentum, according to the meaning of the generalized co-ordinates. Since the instantaneous state of a system is uniquely characterized by giving the position and velocity (momentum) of every point, we speak of momentum *co-ordinates* instead of momentum components, as the  $\dot{q}_k$  and the  $q_k$  have already been given equal rank in the Lagrangian equations of the second kind.

According to equation (43), the momentum co-ordinates are linear functions of the generalized velocity components  $\dot{q}_k$ . Since the determinant of the system of equations does not vanish (as a more thorough analysis shows) the system may be solved for the  $\dot{q}_k$ , so that, reciprocally, the velocity components  $\dot{q}_k$  are also linear functions of the generalized momentum co-ordinates  $p_k$ . If we express the equations of motion in terms of the generalized momenta, we obtain a new form, due to *Hamilton*. This form furnishes the best starting-point



for many investigations in higher mechanics, especially in celestial mechanics, and so has come to be called the *canonical form* of the equations of motion. Introduction of the momentum co-ordinates (43) into the Lagrangian equation (36) (p. 120) gives at once for conservative forces, since  $U$  is independent of the  $\dot{q}_k$ ,

$$\frac{dp_k}{dt} = \frac{\partial L(q_k, \dot{q}_k)}{\partial q_k} \quad . \quad . \quad . \quad . \quad . \quad (44)$$

This equation is still unsatisfactory in that the Lagrangian function in the right member contains the  $\dot{q}_k$  instead of the  $p_k$ . We could try to express the  $\dot{q}_k$  in terms of  $p_k$  by means of the transformation (43), but it is simpler to introduce a new function  $H$  of  $q_k$  and  $p_k$  defined by

$$H(p_k, q_k) = \sum_k p_k \dot{q}_k - L(q_k, \dot{q}_k). \quad . \quad . \quad . \quad . \quad (45)$$

The total differential of this function is

$$\begin{aligned} dH &= \sum_k \left( \frac{\partial H}{\partial p_k} dp_k + \frac{\partial H}{\partial q_k} dq_k \right) \\ &= \sum_k p_k d\dot{q}_k + \sum_k \dot{q}_k dp_k - \sum_k \frac{\partial L}{\partial \dot{q}_k} d\dot{q}_k - \sum_k \frac{\partial L}{\partial q_k} dq_k. \end{aligned} \quad . \quad (46)$$

Since the potential energy depends only upon the position co-ordinates, and not upon those of the velocity, we have

$$\frac{\partial L}{\partial \dot{q}_k} = \frac{\partial T(q_k, \dot{q}_k)}{\partial \dot{q}_k} = p_k. \quad . \quad . \quad . \quad . \quad (47)$$

As a result, the first and the third terms in the right member of equation (46) drop out, and comparison of coefficients gives

$$\dot{q}_k = \frac{\partial H}{\partial p_k}, \quad . \quad . \quad . \quad . \quad . \quad (48)$$

and

$$\frac{\partial L}{\partial q_k} = - \frac{\partial H}{\partial q_k}. \quad . \quad . \quad . \quad . \quad (49)$$

Substituting these results in the "mixed" equation (44), we obtain

$$\frac{dp_k}{dt} = - \frac{\partial H}{\partial q_k} \quad . \quad . \quad . \quad . \quad . \quad (49')$$

The canonical form of the equations of motion is thus

$$\left. \begin{aligned} \frac{dp_k}{dt} &= - \frac{\partial H(p_k, q_k)}{\partial q_k} \\ \frac{dq_k}{dt} &= \frac{\partial H(p_k, q_k)}{\partial p_k} \end{aligned} \right\} \quad . \quad . \quad . \quad . \quad . \quad (50)$$

For conservative forces, to which we restrict ourselves, the meaning of  $H$  is very simple. If we form the total time derivative

$$\frac{dH}{dt} = \sum_k \frac{\partial H}{\partial q_k} \frac{dq_k}{dt} + \sum_k \frac{\partial H}{\partial p_k} \frac{dp_k}{dt}, \quad . . . \quad (51)$$

this derivative vanishes, on account of (50).  $H$  is thus a constant of the system. Using equation (45) (p. 122), it is readily shown that  $H$  is the same as the total energy  $E$ : the kinetic energy  $T$  is a homogeneous quadratic function of the  $\dot{q}_k$ , according to equation (41) (p. 121). According to Euler's Theorem on homogeneous functions, we therefore have

$$\sum_k \dot{q}_k \frac{\partial T}{\partial \dot{q}_k} = 2T = \sum_k \dot{q}_k p_k, \quad . . . \quad (52)$$

as may be verified readily in this instance. It then follows from equation (45) that

$$H(p_k, q_k) = 2T - (T - U) = T + U = E. \quad . \quad (53)$$

## 7. Hamilton's Principle

Besides the Lagrangian and Hamiltonian forms of the equations of motion of a system in arbitrary co-ordinates, there is a third form which is of the greatest importance. This third statement gives the equations not as differential equations, but in the form of a stationary condition. The idea of expressing a law of nature by specifying that a certain quantity is to have an extreme value in the actual process, is as old as scientific thought itself. The advantage lies in the simplicity of the formulation and its independence of any assumptions concerning the particular co-ordinate system selected. The derivation of such an extremal law of mechanics, from what has preceded, is not difficult. Comparison of the Lagrangian equations of the second kind with the Euler differential equation of the simplest variation problem (p. 77) shows that they are identical. In place of the independent variable  $x$  we now have the time, and the function designated by  $F$  on p. 75 is the Lagrangian function  $L$ . Thus the Lagrangian equations of the second kind correspond to an extreme value of the integral

$$W = \int_{t_0}^{t_1} L dt = \int_{t_0}^{t_1} (T - U) dt.$$

The time integral of twice the kinetic energy is called the *action*. We are thus able to express the equations of motion in this manner: *The natural motion of a system (i.e. the actual motion which takes place according to the laws of mechanics) is characterized by the fact that the time*

integral of the Lagrangian function, taken between two configurations of the system has an extreme value. This is the celebrated Hamiltonian Principle. The foregoing derivation offers no information as to whether the value spoken of is a maximum or a minimum, or merely stationary, but this question is of no consequence in what follows. In most instances  $W$  is a minimum.

The mathematical expression of the principle is

$$\delta W = \delta \int_{t_0}^{t_1} L dt = \delta \int_{t_0}^{t_1} (T - U) dt = \delta \int_{t_0}^{t_1} (2T - E) dt = 0, \quad (54)$$

where  $E$  represents the total energy.

If forces which are not derivable from potentials also act, then, by § 5 (p. 120) the Lagrange equations become

$$\frac{d}{dt} \frac{\partial L}{\partial \dot{q}_k} - \frac{\partial L}{\partial q_k} - F'_{qk} = 0. \quad . . . . . (55)$$

In this case the Lagrange equations no longer represent the Euler differential equations of a variation problem. However, they may be brought into this form if it is possible to find a function  $M$  of the co-ordinates and their derivatives such that

$$\frac{d}{dt} \frac{\partial M}{\partial \dot{q}_k} - \frac{\partial M}{\partial q_k} = F'_{qk}. \quad . . . . . (56)$$

In this case the function  $L'$  defined by

$$L' = L - M = T - U - M \quad . . . . . (57)$$

satisfies the differential equation

$$\frac{d}{dt} \frac{\partial L'}{\partial \dot{q}_k} - \frac{\partial L'}{\partial q_k} = 0, \quad . . . . . (58)$$

and this is the Euler differential equation corresponding to

$$\delta \int L' dt = 0. \quad . . . . . (59)$$

This case occurs, for example, in the mechanics of an electron in a magnetic field. The magnetic force (cf. p. 428) is given by  $\mathbf{F}_{\text{mag}} = -(e/c)[\mathbf{v}\mathbf{H}]$ . Since this force is always normal to the path (the tangent to the orbit is given by  $\mathbf{v}$ ) no work is done. It is therefore not possible to derive this force from a scalar potential. If, however, we put  $\mathbf{H} = \text{curl } \mathbf{A}$ , it may be shown that the function  $M = e/c(\mathbf{A}\mathbf{v})$  satisfies (56). Naturally, it is not always possible to find a function  $M$ , i.e. the system of equations (56) may not be integrable.

## 8.\* Canonical Transformations.\*

It is often necessary to change from one system of variables to another. We might proceed by introducing new position co-ordinates by means of equations of the form

\* Sections marked with an asterisk may be omitted at a first reading.



$$q_k = q_k(\bar{q}_1, \bar{q}_2, \dots, \bar{q}_f),$$

and then, using the time derivatives of these equations, express the kinetic energy as a function of the new position co-ordinates and their time derivatives. Finally, with the aid of the equations

$$\bar{p}_k = \frac{\partial T}{\partial \dot{\bar{q}}_k},$$

the new momentum co-ordinates could be determined. Hamilton's principle, however, makes it possible to solve the problem much more elegantly, and at the same time much more generally. Let us change position and momentum co-ordinates *simultaneously*, and inquire how the transformation relationships must be constituted in order that the new variables may again have the character of position and momentum co-ordinates, i.e. that the canonical equations may still be satisfied. For this purpose, it is necessary and sufficient that the new variables again satisfy the Hamiltonian Principle. But, in order that this be true, it is sufficient that the difference of the integrands of the action, expressed in the new and in the old variables respectively, should be the total time derivative of a function of the old and new co-ordinates; that is, by (45) (p. 122),

$$\Sigma p_k \dot{q}_k - H(p_k, q_k) - \Sigma \bar{p}_k \dot{\bar{q}}_k + \bar{H}(\bar{p}_k, \bar{q}_k) = \frac{d}{dt} f(q_k, \bar{q}_k), \quad (60)$$

for if this equation is integrated between the limits  $t_0$  and  $t_1$ , we obtain

$$\begin{aligned} \int_{t_0}^{t_1} [\Sigma p_k \dot{q}_k - H(p_k, q_k)] dt - \int_{t_0}^{t_1} [\Sigma \bar{p}_k \dot{\bar{q}}_k - \bar{H}(\bar{p}_k, \bar{q}_k)] dt \\ = [f(q_k, \bar{q}_k)]_{t_1} - [f(q_k, \bar{q}_k)]_{t_0}. \end{aligned}$$

Upon taking the variation of this expression, the right side vanishes, since the end-points are fixed (cf. p. 78). Since the variation of the first integral is zero, the variation of the second integral must also vanish, which means that the new variables are again canonically conjugate position and momentum co-ordinates. By carrying out the differentiation in the right member of (60), and comparing coefficients, the transformation formulæ are obtained in the form

$$\left. \begin{aligned} p_k &= \frac{\partial f(q_k, \bar{q}_k)}{\partial q_k} \\ \bar{p}_k &= - \frac{\partial f(q_k, \bar{q}_k)}{\partial \bar{q}_k} \\ H(p_k, q_k) &= \bar{H}(\bar{p}_k, \bar{q}_k) \end{aligned} \right\} \dots \dots \dots (61)$$

The  $q_k$  may be computed in terms of the  $\bar{p}_k$  and  $\bar{q}_k$  from the equations

of the second line in (61). If we then substitute these values of the  $q_k$  in the equations of the first line, we obtain the  $p_k$  as functions of the  $\bar{p}_k$  and  $\bar{q}_k$ . These cumbersome calculations are, however, necessary only in rare cases in this general form. The arbitrary function  $f(\bar{q}_k, \bar{p}_k)$  is called the generator of the transformation.

Instead of starting with a function of the old and new position co-ordinates, we may start with a function of old position co-ordinates and new momentum co-ordinates; this leads to relationships of particular importance. Let us add the expression

$$- \frac{d}{dt} \Sigma \bar{p}_k \bar{q}_k$$

to the right member of equation (60). This may be done, since the above expression vanishes when we take the variation of the time integral, just as does the expression  $df/dt$ . We then write

$$\begin{aligned} \Sigma p_k \dot{q}_k - H(p_k, q_k) - \Sigma \bar{p}_k \dot{\bar{q}}_k + \bar{H}(\bar{p}_k, \bar{q}_k) \\ = \frac{df(q_k, \bar{p}_k)}{dt} - \frac{d \Sigma \bar{p}_k \bar{q}_k}{dt} \quad . \quad . \quad (62) \end{aligned}$$

and obtain, by comparison of coefficients,

$$\left. \begin{aligned} p_k &= \frac{\partial f(q_k, \bar{p}_k)}{\partial q_k} \\ \bar{q}_k &= \frac{\partial f(q_k, \bar{p}_k)}{\partial \bar{p}_k} \\ H(p_k, q_k) &= \bar{H}(\bar{p}_k, \bar{q}_k) \end{aligned} \right\} . \quad . \quad . \quad . \quad . \quad (63)$$

### 9.\* Cyclic Variables. The Hamilton-Jacobi Differential Equation.

If the Hamiltonian function—which represents the total energy if the forces are conservative—does not contain a certain position co-ordinate  $q_i$ , then it follows from the first of the Hamiltonian equations that

$$\frac{dp_i}{dt} = - \frac{\partial H}{\partial q_i} = 0. \quad . \quad . \quad . \quad . \quad . \quad (64)$$

That is, the conjugate momentum co-ordinate is a constant of the system. Such a position co-ordinate is called *cyclic*. The name is due to the fact that the polar angle  $\phi$  in central motion has this property. If all the position co-ordinates are cyclic, it follows from the second of Hamilton's equations, viz.

$$\frac{dq_k}{dt} = \frac{\partial H(p_1, p_2, \dots, p_r)}{\partial p_k} = f_k(p_1, p_2, \dots, p_r), \quad . \quad (65)$$

that the  $dq_k/dt$ , being given functions of constants of the system, are themselves constants, and the position co-ordinates increase linearly with time. The integration of the equations of motion may thus be performed immediately in the case of cyclic variables. A very important method for the integration of the equations of motion is based on this fact. Instead of starting with Lagrange's or Hamilton's equations, for example, and trying to integrate them, *we can begin by trying to find the generator of a canonical transformation which leads to only cyclic variables*. If we call the momentum co-ordinates, which are constant in time,  $\alpha_k$ , this generator must be a function  $S$  of the old position co-ordinates and of the new momentum co-ordinates, such that

$$H(p_k, q_k) = \bar{H}(\alpha_k). \quad . \quad . \quad . \quad . \quad . \quad (66)$$

According to equation (63) we must put

$$p_k = \frac{\partial S(q_k, \alpha_k)}{\partial q_k} \quad \text{and} \quad \bar{q}_k = \phi_k = \frac{\partial S}{\partial \alpha_k}. \quad . \quad . \quad (67)$$

The function  $S$  is, then, determined by the following differential equation: if we substitute  $\partial S/\partial q_k$  for  $p_k$  in the Hamiltonian function, we have for conservative systems

$$H\left(\frac{\partial S}{\partial q_k}, q_k\right) = E, \quad . \quad . \quad . \quad . \quad . \quad (68)$$

which is the *Hamilton-Jacobi Differential Equation*. Here  $E$  is the energy of the system, and must thus be considered a constant of the system, since the differential equations themselves do not contain the *value* of the energy. We are therefore at liberty to identify  $E$  with one of the constants  $\alpha_k$ , e.g.  $\alpha_1$ . The Hamilton-Jacobi differential equation is of the first order, but second degree, since  $H$  is a quadratic function of the  $p_k$ ; the complete solution, then, contains  $f$  constants of integration (if there are  $f$  position co-ordinates), one of which may be set equal to  $E$ . Instead, we could introduce  $f$  arbitrary constants; but in this case  $E$  must be a function of the  $\alpha_1, \alpha_2, \dots, \alpha_f$ , and may no longer be taken as an arbitrary integration constant. Thus, the solution is either

$$S(q_k, E, \alpha_2, \alpha_3 \dots \alpha_f) = 0; \quad E = \alpha_1, \quad . \quad . \quad . \quad (69)$$

or more generally,

$$S(q_k, \alpha_1, \alpha_2, \alpha_3 \dots \alpha_f) = 0; \quad E = E(\alpha_1, \alpha_2 \dots \alpha_f). \quad (70)$$



In the former case, we see by (65) that all of the  $d\bar{q}_k/dt$  with the exception of  $d\bar{q}_1/dt$  are zero, since \*  $E = H$ .

There is an intimate connexion between the function  $S$  introduced here as the generator of a transformation to cyclic variables and the action function (Hamilton's characteristic function). During the course of the motion, the  $\alpha_k$  are constant, and so their differentials are zero. This means that the differential of  $S$  is given by

$$dS = \sum \frac{\partial S}{\partial q_k} dq_k = \sum p_k dq_k = \sum p_k \dot{q}_k dt. \quad . \quad . \quad (71)$$

Now by equation (52) (p. 123),

$$\sum p_k \dot{q}_k = 2T,$$

$$\text{i.e.} \quad dS = 2T dt \quad . \quad . \quad . \quad . \quad . \quad . \quad (72)$$

$$\text{and} \quad S - S_0 = 2 \int_{t_0}^{t_1} T dt. \quad . \quad . \quad . \quad . \quad . \quad . \quad (73)$$

On the other hand, the action function is, by § 7 (p. 123),

$$W = \int 2T dt - Et. \quad . \quad . \quad . \quad . \quad . \quad . \quad (74)$$

$$\text{Therefore,} \quad W = S - Et. \quad . \quad . \quad . \quad . \quad . \quad . \quad (75)$$

Thus  $S$  represents the part of the action  $W$  which does not contain the time implicitly. On account of the relationship (75),  $S$  is also known as the contracted action function.

#### 10.\* Periodic and Multiply Periodic Systems. Angle Variables; the Angle Variables of Keplerian Motion.

We consider first a system having a single position co-ordinate  $x$ , e.g. a linear oscillator. The system is said to be periodic with period  $T$ , if the equation

$$x(t + nT) = x(t)$$

is satisfied when  $n$  is any whole number. According to § 9, a cyclic variable  $\phi$  increases linearly with the time, i.e.

$$\phi = \gamma t, \quad t = \frac{\phi}{\gamma}.$$

For this reason, the co-ordinate  $x$  is also a periodic function of the cyclic co-ordinate  $\phi$ . We then call the cyclic co-ordinate  $\phi$  an angle co-ordinate  $w$  if the period is unity, i.e. when

$$x(w + 1) = x(w). \quad . \quad . \quad . \quad . \quad . \quad . \quad (76)$$

\* It is impossible to avoid the double designation of the energy by  $E$  and by  $H$ , for the following reason:  $H(p_k, q_k)$  signifies the Hamiltonian function of the position and momentum co-ordinates. For conservative systems, this is identical with the total energy  $E$ , which is a constant of the system. On passing to cyclic variables  $H$  will depend also only on quantities constant in time. In this case we may write  $E$  in place of  $H$ .

In the simple case of a point moving uniformly around a circle, the polar angle  $\phi$  is a cyclic variable, but the angle variable is the quantity  $w = \phi/2\pi$ , for in the latter quantity the rectangular co-ordinate  $x$  is periodic with period unity. The momentum co-ordinate conjugate to an angle co-ordinate is called an *action variable*, and is designated by  $J$ .

The advantage of using angle and action variables lies in the fact that we obtain the frequency of the system immediately by differentiation of the Hamiltonian function; for if  $H$  is given as a function of the action variables ( $H$  cannot be a function of the conjugate position co-ordinates, since every angle co-ordinate is cyclic), then

$$\frac{dw}{dt} = \frac{dH}{dJ} = \gamma, \quad \text{i.e. } w = \gamma t + \delta. \quad . . . (77)$$

On the other hand, every periodic function of frequency  $\nu$  is representable, subject to conditions as to continuity which are by no means onerous, by a Fourier series of the form

$$x = \sum c_k e^{2\pi i n_k \nu t}. \quad . . . . . (78)$$

The series is a periodic function of the quantity  $\nu t$ , with period unity. But  $w$  is an angle variable by hypothesis, and so  $x$  is also a periodic function of  $\gamma t$ , with period 1, i.e.  $\gamma$  is identical with  $\nu$ . What constant of the system must be introduced as a new momentum co-ordinate in order to make the conjugate position co-ordinate not only a cyclic variable but also an angle variable? At any rate, the generator of the transformation must satisfy the Hamilton-Jacobi differential equation, and is therefore again denoted by  $S$ . If  $S$  is known as a function of  $q$  and  $J$ , then we have, as in § 9,

$$w = \frac{\partial S(q, J)}{\partial J} \quad \text{and} \quad p = \frac{\partial S(q, J)}{\partial q}. \quad . . . (79)$$

On account of the constancy of  $J$ , the increase of  $w$  during the course of the motion is

$$dw = \frac{\partial^2 S(q, J)}{\partial q \partial J} dq. \quad . . . . . (80)$$

Our requirement that  $w$  be an angle variable may be expressed as follows: when the system has returned to its former state,  $w$  must have increased by 1. If we integrate the increase of  $w$  over a period, we must have

$$1 = \oint dw = \oint \frac{\partial^2 S}{\partial q \partial J} dq = \frac{\partial}{\partial J} \oint \frac{\partial S}{\partial q} dq = \frac{\partial}{\partial J} \oint p dq. \quad (81)$$





$x$  increases, so that  $\dot{x}$ , and therefore  $p$  and  $dS/dx$ , are positive. On the return,  $x$  decreases, and  $dS/dx$  has the negative sign. The complete integral

$$J = \oint \frac{dS}{dx} dx = \oint \sqrt{2mE - mkx^2} dx = \sqrt{mk} \oint \sqrt{\frac{2E}{k} - x^2} dx$$

thus represents  $\sqrt{mk}$  times the area of a circle of radius  $\sqrt{2E/k}$ . That is,

$$J = \frac{2E}{k} \pi \sqrt{mk} \quad \text{or} \quad E = \frac{J}{2\pi} \sqrt{\frac{k}{m}}.$$

From this result, the frequency is

$$\nu = \frac{dH}{dJ} = \frac{1}{2\pi} \sqrt{\frac{k}{m}},$$

as already derived on p. 94 by elementary methods. Indefinite integration of  $dS/dx$ , and the replacing of  $E$  by

$$\frac{J}{2\pi} \sqrt{\frac{k}{m}}$$

yields the generator  $S$  of the transformation to angle variables, and differentiation with respect to  $J$  gives  $w$ . The calculation is not troublesome, since the integral

$$\int \sqrt{2mE - mkx^2} dx$$

is easily evaluated as an indefinite integral. The result is

$$w = \frac{1}{2\pi} \sin^{-1} \sqrt{\frac{\pi \sqrt{km}}{J}} x. \quad \dots \quad (85)$$

The Hamilton-Jacobi method appears much more complicated here than the direct integration, as carried out in Chap. V, § 7. As a matter of fact, the full value of the Hamilton-Jacobi method first becomes evident in connexion with complicated systems in Celestial Mechanics, or Atomic Mechanics, where it proves to be an indispensable aid.

The application of these principles to systems of more than one degree of freedom requires development of the concept of multiply periodic—or conditionally periodic—systems, which occur when there are several degrees of freedom. Let the rectangular co-ordinates  $x_k$  be periodic functions of the angle variables  $w_k$  with period 1, i.e.

$$x_k(w_1 + 1, w_2 + 1, w_3 + 1, \dots) = x_k(w_1, w_2, w_3, \dots), \quad (86)$$

and generally, for whole numbers  $n_k$ ,

$$x_k(w_1 + n_1, w_2 + n_2, \dots) = x_k(w_1, w_2, \dots). \quad (87)$$

Just as Fourier's series is the most general representation of periodic

functions of one variable, so the general analytical expression for functions of several variables, having as many periods as there are variables, is the multiple Fourier series of the form

$$x_k = \sum_l \sum_m \sum_p \dots C_{lm p \dots}^k e^{2\pi i(n_l w_1 + n_m w_2 + n_p w_3 + \dots)}. \quad (88)$$

Since the  $w_k$  increase linearly with the time, we may substitute the expressions  $\nu_k t + \delta_k$  for  $w_k$  in the Fourier series. Taking the factors  $e^{2\pi i n_k \delta_k}$  originating in the phase constants into the coefficients, we then have

$$x_k = \sum_l \sum_m \sum_p \dots D_{lm p \dots}^k e^{2\pi i(n_l \nu_1 t + n_m \nu_2 t + n_p \nu_3 t + \dots)}. \quad (89)$$

This multiple Fourier series no longer represents a periodic function of the time. It is true that any one factor, e.g.  $e^{2\pi i n_l \nu_1 t}$  returns to its initial value after  $T_1 = 1/\nu_1$  sec; but this term is multiplied by other factors  $e^{2\pi i n_m \nu_k t}$  which certainly do not return to their old values since the separate  $\nu_k$  are taken to be independent of each other. A true periodicity in time can be obtained only by imposing additional conditions on the  $\nu_k$ , e.g. the condition  $\nu_1 = \nu_2 = \nu_3 = \nu_k$ .

For this reason, multiply periodic systems are also called *conditionally periodic*. A simple example of a conditionally periodic system is a particle performing harmonic vibrations along the  $x$ - and  $y$ -axes, the two vibrations having incommensurable frequencies. As shown on p. 60, the resulting orbit is a so-called Lissajous figure which never closes, if the ratio of the two frequencies is not a rational number, but which, in the course of time, covers all points within a given rectangle. All orbits of conditionally periodic motions are of the type of the Lissajous figure; a certain portion of space is densely filled, yet the orbit is not closed, and thus a true periodicity in time does not exist.

Conditionally periodic motions occur in general if the Hamilton-Jacobi equation (68) (p. 127) can be solved by "separation of the variables"—the only feasible practical method, assuming that the system has properties of periodicity at all. By separation of the variables we mean the solution of the equation by putting, in those cases where it is possible,

$$S(q_1, q_2, q_3 \dots q_f) = S_1(q_1) + S_2(q_2) + S_3(q_3) + \dots S_f(q_f), \quad (90)$$

so that

$$p_k = \frac{\partial S}{\partial q_k} = \frac{dS_k}{dq_k}, \quad \dots \dots \dots (91)$$

and the equation may be separated into a number of ordinary differential equations of the form

$$\left(\frac{dS_k}{dq_k}\right)^2 + f(q_k) = a_k, \quad \dots \dots \dots (92)$$

the energy  $E$  becoming a function of the constants of integration:

$$E = E(a_1, a_2, a_3, \dots a_f). \quad \dots \dots \dots (93)$$

The search for the angle and action variables proceeds just as in the case of one degree of freedom treated above: let the angle and action variables which are sought be denoted by  $w_k$  and  $J_k$  respectively, and let  $S$  be the generator of the transformation, as a function of the old position co-ordinates  $q_k$  and the new momentum co-ordinates  $J_k$ .

Imagine all the position co-ordinates  $q_k$  except  $q_h$  held constant. Since the action variables are constants, the change in the angle variables corresponding to a change in  $q_h$  is

$$d_h w_k = \frac{\partial w_k}{\partial q_h} dq_h. \quad \dots \dots \dots (94)$$

Now, since

$$w_k = \frac{\partial S(q_k, J_k)}{\partial J_k},$$

the above expression becomes

$$d_h w_k = \frac{\partial^2 S}{\partial q_h \partial J_k} dq_h. \quad \dots \dots \dots (95)$$

We now postulate the separability of the Hamilton-Jacobi equation in the co-ordinates  $q_k$ , which we therefore call the separation variables. Thus

$$d_h w_k = \frac{\partial}{\partial J_k} \frac{dS_h}{dq_h} dq_h. \quad \dots \dots \dots (96)$$

The change in  $w_k$  corresponding to passage through the complete range of variation of  $q_h$  is

$$\Delta_h w_k = \frac{\partial}{\partial J_k} \oint \frac{dS_h}{dq_h} dq_h. \quad \dots \dots \dots (97)$$

If we now introduce the phase integrals

$$J_h = \oint \frac{dS_h}{dq_h} dq_h = \oint p_h dq_h \quad \dots \dots \dots (98)$$

as action variables, then

$$\left. \begin{aligned} \Delta_h w_k &= 1 \text{ for } h = k \\ &= 0 \text{ for } h \neq k \end{aligned} \right\} \dots \dots \dots (99)$$

But this means that the variable of separation  $q_k$  is a periodic function of  $w_k$ , with period unity. While it is true that each of the variables  $q_k$  is simply periodic, the periods of the separate  $q_k$  are different, and so the system is multiply periodic. Hence the rectangular co-ordinates,



which may depend on the  $q_k$  in many ways, become multiply periodic functions of the  $w_k$ . The variables  $w_k$  and  $J_k$ , introduced in this manner, thus fulfil all the conditions imposed above on angle and action variables.

If the energy is expressed in terms of the phase integrals rather than the constants of integration  $\alpha_k$ , the frequencies of the system follow by differentiation since, according to (65) (p. 126), and the above discussion,

$$\nu_k = \frac{dw_k}{dt} = \frac{\partial H}{\partial J_k}.$$

Let us illustrate the method with the example of Keplerian motion treated on p. 90. Since the force is a function only of the distance  $r$  from the centre of attraction, three-dimensional polar co-ordinates are suggested. The line element is given by

$$ds^2 = dr^2 + r^2 d\theta^2 + r^2 \sin^2 \theta d\phi^2. \quad . \quad . \quad . \quad (100)$$

Then 
$$T = \frac{1}{2}m \left( \frac{ds}{dt} \right)^2 = \frac{1}{2}m(\dot{r}^2 + r^2 \dot{\theta}^2 + r^2 \sin^2 \theta \dot{\phi}^2), \quad . \quad . \quad (101)$$

and the momentum co-ordinates become

$$p_r = m\dot{r}, \quad p_\theta = mr^2\dot{\theta}, \quad p_\phi = mr^2 \sin^2 \theta \dot{\phi}. \quad . \quad . \quad . \quad (102)$$

Expressed in terms of these co-ordinates, the kinetic energy becomes

$$T = \frac{1}{2m} \left( p_r^2 + \frac{1}{r^2} p_\theta^2 + \frac{1}{r^2 \sin^2 \theta} p_\phi^2 \right). \quad . \quad . \quad . \quad (103)$$

The potential energy for the gravitational force between  $m$  and the central mass  $M$  is

$$U = -\frac{\gamma m M}{r} = -\frac{b}{r}. \quad . \quad . \quad . \quad . \quad (104)$$

Replacing the  $p_k$  by  $\partial S / \partial q_k$  in the Hamilton-Jacobi differential equation, we have

$$\frac{1}{2m} \left[ \left( \frac{\partial S}{\partial r} \right)^2 + \frac{1}{r^2} \left( \frac{\partial S}{\partial \theta} \right)^2 + \frac{1}{r^2 \sin^2 \theta} \left( \frac{\partial S}{\partial \phi} \right)^2 \right] - \frac{b}{r} = E. \quad . \quad (105)$$

We put

$$S = S_r(r) + S_\theta(\theta) + S_\phi(\phi), \quad . \quad . \quad . \quad . \quad (106)$$

and obtain

$$\frac{1}{2m} \left[ \left( \frac{dS_r}{dr} \right)^2 + \frac{1}{r^2} \left( \frac{dS_\theta}{d\theta} \right)^2 + \frac{1}{r^2 \sin^2 \theta} \left( \frac{dS_\phi}{d\phi} \right)^2 \right] - \frac{b}{r} = E. \quad . \quad (107)$$

We can now separate the equation into three equations with  $a_\phi$  and  $a_\theta^2$  as constants of integration:

$$\frac{dS_\phi}{d\phi} = a_\phi, \quad . \quad . \quad . \quad . \quad (108)$$

$$p_\phi = \frac{dS_\phi}{d\phi} = a_\phi, \quad . \quad . \quad . \quad . \quad (108')$$

$$\left( \frac{dS_\theta}{d\theta} \right)^2 + \frac{a_\phi^2}{\sin^2 \theta} = a_\theta^2, \quad . \quad . \quad . \quad . \quad (109)$$

$$\text{or } p_\theta = \frac{dS_\theta}{d\theta} = \sqrt{a_\theta^2 - \frac{a_\phi^2}{\sin^2 \theta}}, \quad \dots \dots \dots (109')$$

$$\frac{1}{2m} \left[ \left( \frac{dS_r}{dr} \right)^2 + \frac{a_\theta^2}{r^2} \right] - \frac{b}{r} = E, \quad \dots \dots \dots (110)$$

$$p_r = \frac{dS_r}{dr} = \sqrt{2mE + \frac{2mb}{r} - \frac{a_\theta^2}{r^2}}. \quad \dots \dots \dots (110')$$

These equations are somewhat more general than (92) and (93), since (109) and (110) contain the integration constants of (108) and (109) respectively. The subsequent procedure, however, is the same as above. By the second form, the integration is reduced to quadratures. In order to calculate the phase integrals  $J_\theta$ , the limits of the variables must be determined; the second form of the equations is adequate for this purpose. Since  $a_\phi$  is constant,  $p_\phi$ , and therefore  $\phi$ , always have the same algebraic sign. If the system is conditionally periodic at all, then a complete period of  $\phi$  is from 0 to  $2\pi$ . Then

$$J_\phi = \int_0^{2\pi} a_\phi d\phi = 2\pi a_\phi; \quad a_\phi = \frac{1}{2\pi} J_\phi. \quad \dots \dots \dots (111)$$

The co-latitude  $\theta$  behaves differently. Since  $p_\theta$  must remain real and finite, the limits are given by

$$\sin \theta_{\max} = \sin(\pi - \theta_{\min}) = \frac{a_\phi}{a_\theta}. \quad \dots \dots \dots (112)$$

With increasing  $\theta$ , the positive sign of the radical is to be taken; for decreasing  $\theta$ , the negative sign, as explained in detail for the case of the linear oscillator (p. 130). The value of the integral taken from  $\theta_{\min}$  to  $\theta_{\max}$  and back to  $\theta_{\min}$  is

$$J_\theta = \oint \sqrt{a_\theta^2 - \frac{a_\phi^2}{\sin^2 \theta}} d\theta = 2\pi(a_\theta - a_\phi), \quad \dots \dots \dots (113)$$

as may be calculated readily. Then

$$a_\theta = \frac{J_\theta + J_\phi}{2\pi}.$$

If we substitute this value of  $a_\theta$  in the expression for  $dS_r/dr$ , we obtain

$$J_r = \oint \sqrt{2mE + \frac{2mb}{r} - \frac{(J_\theta + J_\phi)^2}{4\pi^2 r^2}} dr. \quad \dots \dots \dots (114)$$

If the motion is periodic or conditionally periodic, then, here too, the zeros of the expression under the sign of the square root, or radicand, determine a minimum and a maximum of  $r$ , and these quantities must be positive, since the radius vector is always positive. The solution of the quadratic in  $r$ , obtained by setting the radicand equal to zero, gives two positive values of  $r$  only if  $E$  is negative. This is in harmony with the result of Chap. V, § 6, p. 92, that the orbit is hyperbolic for  $E$  positive—in which case there is no finite extreme value of  $r$ —while negative total energy leads to the periodic elliptic orbit. With the assumption that  $E$  is negative, the integral is to be extended from  $r_{\min}$  to  $r_{\max}$ , and back to  $r_{\min}$  with the opposite sign for the radical. The evaluation of the integral by elementary means is somewhat cumbersome, but possible; however, the result is obtained much more readily by complex integration (of. p. 73), and is

$$J_r = -J_\theta - J_\phi + \frac{2\pi mb}{\sqrt{-2mE}}. \quad \dots \dots \dots (115)$$

From this we have

$$E = \frac{-2\pi^2 mb^2}{(J_r + J_\theta + J_\phi)^2}, \quad \cdot \cdot \cdot \cdot \cdot \cdot (116)$$

which gives the energy  $E$  in terms of the action variables:

$$E = H(J_r, J_\theta, J_\phi). \quad \cdot \cdot \cdot \cdot \cdot \cdot (117)$$

The frequencies may be determined at once by differentiation. Here we have the special case where

$$\nu_r = \frac{\partial H}{\partial J_r} = \nu_\theta = \frac{\partial H}{\partial J_\theta} = \nu_\phi = \frac{\partial H}{\partial J_\phi}. \quad \cdot \cdot \cdot \cdot (118)$$

since the energy depends only upon the sum  $J_r + J_\theta + J_\phi$ . This is in keeping with the fact that the elliptic orbit is closed, and hence that we are dealing with a true time-periodicity.

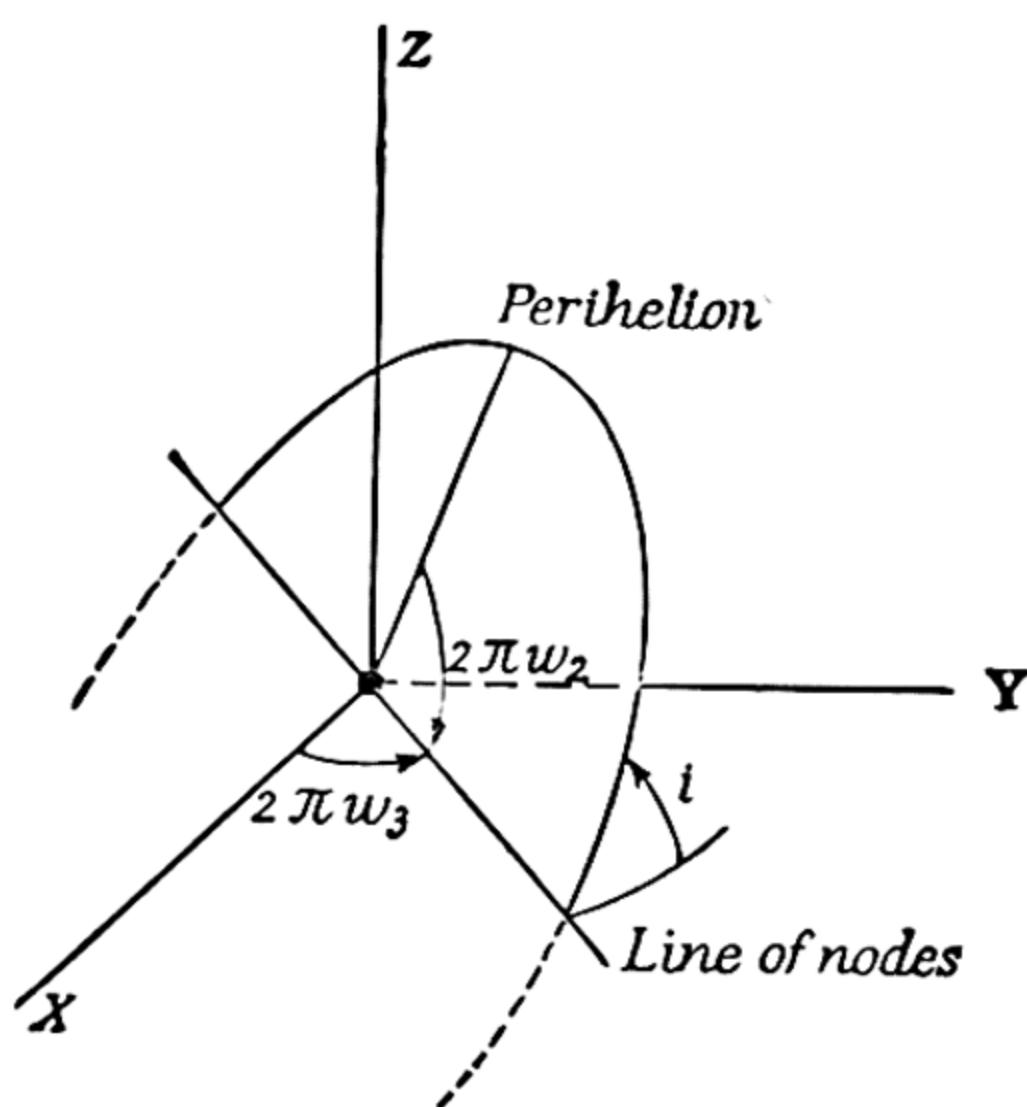


Fig. 1

Since we may set up as many phase integrals as there are co-ordinates, and since the energy, in general, depends upon these integrals individually, one should expect as many independent frequencies as there are degrees of freedom—in the present case, three. If the number of independent frequencies is smaller, the system is said to be *degenerate*. The difference between the number of degrees of freedom and the number of frequencies is called the *degree of degeneracy*. The Keplerian motion is thus doubly degenerate. In this very frequently occurring case of degeneracy, the quantities  $J_r, J_\theta, J_\phi$  are not yet the correct action variables. If we remove the degeneracy by introducing additional small forces ("perturbations"), we must insist that some of the angle variables of the resulting non-degenerate system pass over into constants of the degenerate system when the perturbing forces are reduced to zero. This is not the case, however, for the angle variables which correspond to  $J_r, J_\theta, J_\phi$ . The true angle variables may be obtained by starting with a non-degenerate system and allowing the perturbing forces to vanish. These variables, which are known in astronomy as the *Delaunay elements*, have come to be of great importance. We quote them here without verification: the first action variable is the sum

$$J_1 = J_r + J_\theta + J_\phi.$$



The associated angle variable is the *mean anomaly*  $M$ , which is related to the *eccentric anomaly*  $u$ , occurring in the parametric representation of the ellipse  $x = a \cos u$ ,  $y = b \sin u$ , by the relationship

$$M = \frac{1}{2\pi} (u - \varepsilon \sin u),$$

where  $\varepsilon$  is the eccentricity. The second action variable is  $2\pi$  times the total angular momentum:  $J_2 = 2\pi P$ . The associated position co-ordinate is given by  $1/2\pi$  times the angle between perihelion and the *line of nodes*, the latter being the line of intersection of the orbital plane and the  $xy$ -plane of a fixed co-ordinate system. This co-ordinate is constant for an unperturbed Kepler ellipse. The third momentum co-ordinate  $J_3$  is  $2\pi$  times the  $z$ -component of the angular momentum, i.e.  $J_3 = 2\pi P_z$ . The associated position co-ordinate  $w_3$  is  $1/2\pi$  times the angle between the line of nodes and the  $x$ -axis; this also remains fixed for an unperturbed Kepler ellipse (fig. 1).

## CHAPTER VII

### THE MECHANICS OF RIGID BODIES

#### 1. Selected Topics in the Kinematics of Rigid Bodies.

##### (a) *The number of degrees of freedom of a rigid body*

A rigid body is understood to be a system of particles, the individuals of which are rigidly inter-connected, i.e. always maintain the same distance from each other. We therefore leave out of consideration deformations which occur in all actual physical bodies.

The position of a rigid body with respect to an arbitrary co-ordinate system fixed in space is determined if the position of the origin, and the orientation, of a co-ordinate system fixed in the body are given. In order to fix the origin of the latter co-ordinate system we need three co-ordinates; to specify the orientation of this system, we may use the direction cosines of the axes referred to the system fixed in space. If we denote these cosines by  $\alpha_i, \beta_i, \gamma_i$ , then

$$\left. \begin{aligned} i' &= \alpha_1 i + \beta_1 j + \gamma_1 k \\ j' &= \alpha_2 i + \beta_2 j + \gamma_2 k \\ k' &= \alpha_3 i + \beta_3 j + \gamma_3 k \end{aligned} \right\} \cdot \cdot \cdot \cdot \cdot \quad (1)$$

But there are six equations connecting these nine coefficients, viz.:

$$\left. \begin{aligned} 1 &= i'^2 = \alpha_1^2 + \beta_1^2 + \gamma_1^2 \\ 1 &= j'^2 = \alpha_2^2 + \beta_2^2 + \gamma_2^2 \\ 1 &= k'^2 = \alpha_3^2 + \beta_3^2 + \gamma_3^2 \end{aligned} \right\}, \cdot \cdot \cdot \cdot \cdot \quad (2)$$

$$\left. \begin{aligned} 0 &= i'j' = \alpha_1\alpha_2 + \beta_1\beta_2 + \gamma_1\gamma_2 \\ 0 &= j'k' = \alpha_2\alpha_3 + \beta_2\beta_3 + \gamma_2\gamma_3 \\ 0 &= k'i' = \alpha_3\alpha_1 + \beta_3\beta_1 + \gamma_3\gamma_1 \end{aligned} \right\}, \cdot \cdot \cdot \cdot \cdot \quad (3)$$

i.e. there are but three additional free variables. The position of a rigid body is thus completely determined by six numbers (co-ordinates); *the number of degrees of freedom is therefore six.*

It often proves useful to introduce three independent variables, in place of the nine direction cosines conditioned by six equations. For this purpose the Euler angles  $\theta, \psi, \phi$  are suitable. These angles

are defined as follows (cf. fig. 1):  $\theta$  is the angle between the  $z'$ -axis and the  $z$ -axis. The  $xy$ -plane cuts the  $x'y'$ -plane in a straight line, which we again call the line of nodes, and which we define by the unit vector  $n$ . We draw a straight line and corresponding unit vector  $m'$  perpendicular to  $n$  and in the  $x'y'$ -plane. The angle  $\psi$  is the longi-

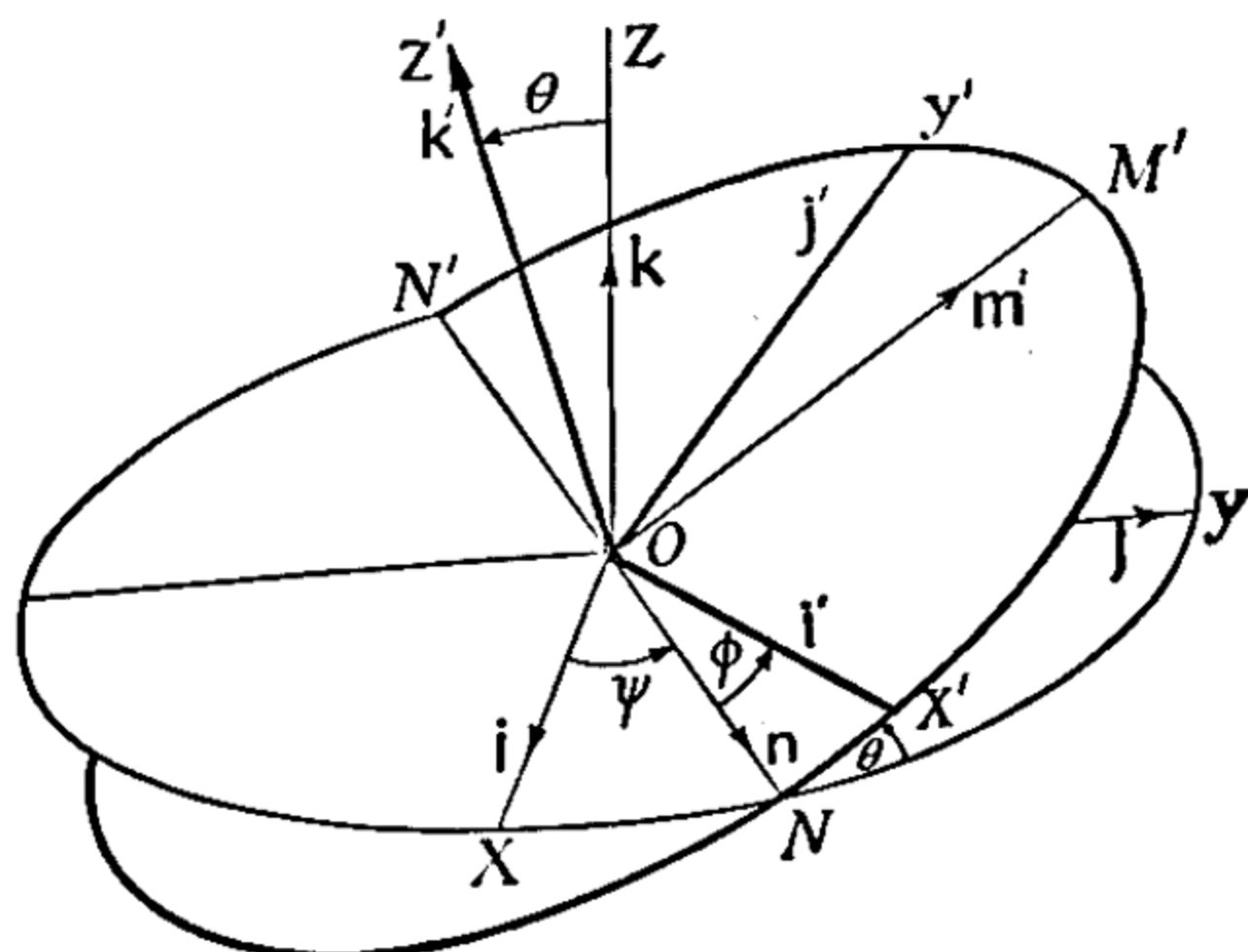


Fig. 1

tude of the node, i.e. the angle between the line of nodes and the  $x$ -axis.  $\phi$  denotes the angle, in the  $x'y'$ -plane, between the  $x'$ -axis and the line of nodes. If the Euler angles are given, it is not difficult to draw the axes  $i', j', k'$ . Using  $\psi$ , one can first draw the line of nodes in the  $xy$ -plane. Through this line pass a plane making an angle  $\theta$  with the  $xy$ -plane; this is the  $x'y'$ -plane. The normal to it is the  $z'$ -axis. In the  $x'y'$ -plane, the position of the  $x'$ -axis is then determined by  $\phi$ .

The axes  $O_{x'}, O_{y'}, O_{z'}$  are fixed in the body, and  $O_x, O_y, O_z$  are fixed in space.

The number of degrees of freedom is diminished if separate points of the rigid body are held fixed. If, for example, one point is fixed, we place the origin of co-ordinates  $O$  at that point; the position of the body is then determined by the three Euler angles  $\theta, \psi, \phi$ , and so the number of degrees of freedom is reduced to three. If we hold two points fixed, the only possible motion is rotation about the axis formed by these points. The position is then fully determined by giving the angle of rotation, i.e. the angle between a line in the body normal to the axis and a plane fixed in space, through that axis. A rigid body in which two points are fixed has then but one degree of freedom. This last degree of freedom is also removed if the position of another point, not in the axis, is prescribed. We shall consider the motions possible for a rigid body under these various limitations, and finally shall touch upon the general motion of a free rigid body.



(b) *Rotation of a rigid body about a fixed axis*

The angular velocity and the direction of the axis together may be characterized by a vector in the following manner: we assign a vector  $\omega$  in the direction of the axis, such that the rotation is clockwise when looking in the direction of  $\omega$ , and such that the magnitude of  $\omega$  is equal to the angular speed  $\omega$ . This representation has a meaning only if we can show that the angular velocities so specified follow the laws of vector analysis, i.e. that two angular velocities around intersecting axes combine to give a third angular velocity whose axis and magnitude are determined by vector addition. This is actually

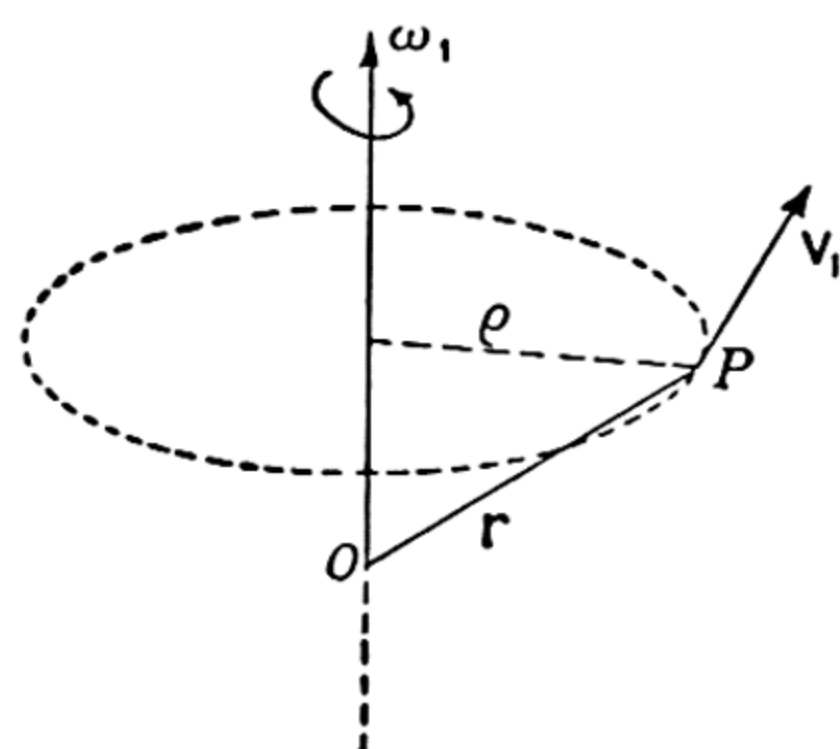


Fig. 2

the case. If  $\omega_1$  is the first angular velocity, the linear velocity  $v_1$  of a point whose radius vector, referred to a point  $O$  on the axis, is  $r$  is given in direction and magnitude by the vector product  $[\omega_1 r]$ ; for the linear speed (see fig. 2) is equal to the product of  $\omega_1$  by the perpendicular distance  $\rho$  of the point  $P$  from the axis, i.e. equal to  $\omega_1 |r| \sin(\omega_1 r)$ . Moreover,  $v_1$  is normal to the axis, which is also true of  $[\omega_1 r]$ . If now a second axis through  $O$  is given, then the linear velocity corresponding to the rotational velocity  $\omega_2$  is given by  $v_2 = [\omega_2 r]$ . The resulting linear velocity is thus

$$v = v_1 + v_2 = [\omega_1 r] + [\omega_2 r],$$

which, by the distributive property of the vector product, is

$$v = |(\omega_1 + \omega_2)r| = |\omega r|. \quad . \quad . \quad . \quad (4)$$

That is, we may replace the angular velocities  $\omega_1$  and  $\omega_2$  by the angular velocity  $\omega$  which is obtained from  $\omega_1$  and  $\omega_2$  by vector addition. If we multiply both members of equation (4) by  $dt$ , we have the same law of combination for infinitesimal angular displacements (rotations).

However, this important rule for combining angular velocities and infinitesimal rotations must not be understood to hold for the analogous case of *finite* rotations about two intersecting axes. True, one may assign to every finite rotation a directed line-segment whose length is proportional to the angle turned, but—as may be seen at once by simple examples—two such rotations combine in a much more complicated manner than two vectors. A finite rotation signifies a transformation of axes, i.e. a tensor; and while two successive transformations may be replaced by a single one, the replacement cannot be made as simply as for angular velocities.

(c) *Plane motion of a rigid body*

Next to rotation about a fixed axis, the simplest case would be the motion of a rigid body about a fixed point, according to the above enumeration of degrees of freedom. However, it is advisable before taking up this case to consider a motion which also has three degrees of freedom—plane motion, in which all points of the body move parallel to a given fixed plane. All points on a line normal to this plane then describe congruent paths, and so it is sufficient to study the motion of one plane moving over another fixed plane. One may picture this motion as that of a sheet of paper moving about on a plane table-top. It is evident that this motion has three degrees of freedom, for if we fix one point of the moving plane by giving its two

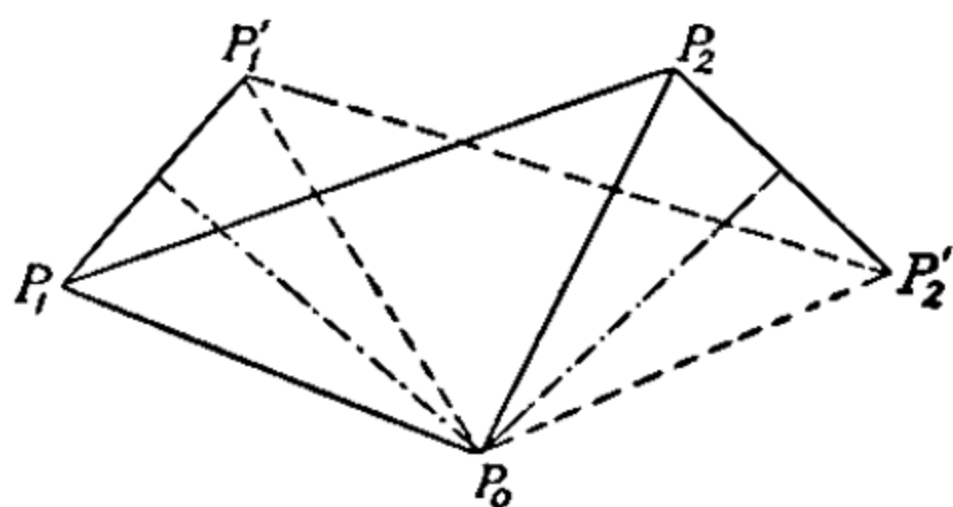


Fig. 3

co-ordinates, we can still turn this plane about the fixed point; the position of the moving plane is completely fixed by further specifying the angle of turn referred to a fixed axis. It may be shown that the general motion at any instant may be looked upon as a rotation about an instantaneous pivot point, the instantaneous centre; for if we select two arbitrary points  $P_1$  and  $P_2$ , the motion of every other point is determined by the motion of these. After a time  $\Delta t$  the two points are at  $P_1'$  and  $P_2'$  respectively (fig. 3). Draw the perpendicular bisectors of  $P_1P_1'$  and  $P_2P_2'$ ; let these intersect at  $P_0$ . The triangles  $P_0P_1P_2$  and  $P_0P_1'P_2'$  are congruent, since the construction makes  $P_0P_1 = P_0P_1'$  and  $P_0P_2 = P_0P_2'$ , and also  $P_1P_2 = P_1'P_2'$  on account of the rigidity of the body. As a result, the triangle  $P_0P_1'P_2'$  can be brought into superposition with triangle  $P_0P_1P_2$  by rotation about  $P_0$ . Since the position of the body is determined, for plane motion, by the two points  $P_1$  and  $P_2$ , which were chosen

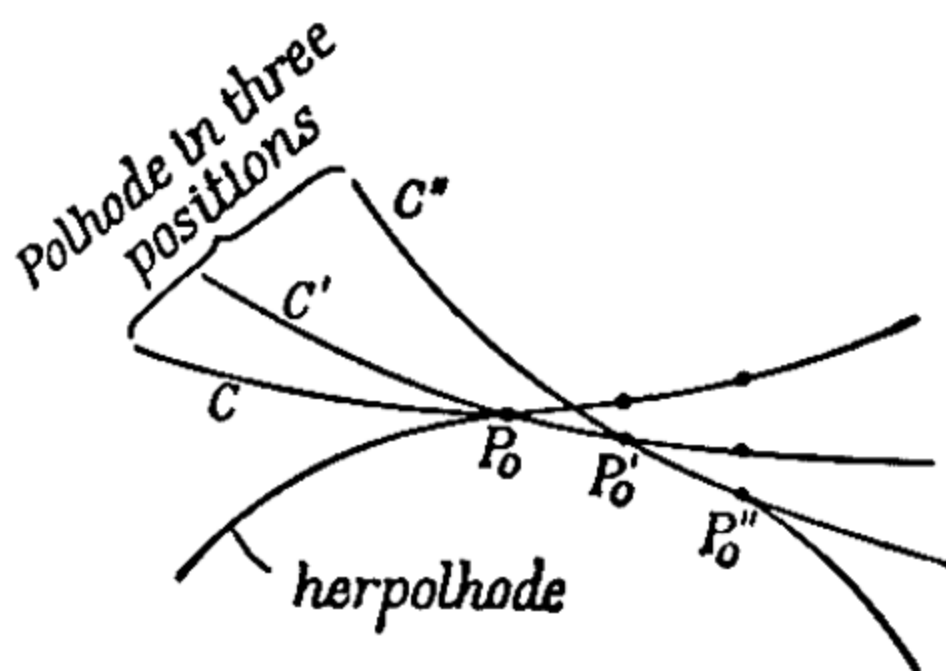


Fig. 4

at random, the same construction with any other pair of points must also lead to the point  $P_0$ . If  $\Delta t$  is allowed to approach zero, the segments  $P_1P_1'$  and  $P_2P_2'$  become the tangents of the paths, and have the directions of the velocity vectors, and are normal to  $P_1P_0$  and  $P_2P_0$  respectively. *The instantaneous state of motion may therefore be looked upon as a rotation about  $P_0$ . This point is the instantaneous centre.*

If the position of the instantaneous centre is marked in the fixed



plane, as well as in the moving plane, we obtain (except in the trivial case of simple rotation about a fixed point) a curve in each plane. The curve in the fixed plane is called the *herpolhode*; that in the moving system, the *polhode*. According to their definition, both curves must have the instantaneous centre in common at any instant. An infinitesimal rotation about this point must bring the neighbouring points of both curves into coincidence. This is possible only if the two curves touch at the pole and roll upon each other without slipping. This is immediately evident from a diagram in which the curves are provisionally replaced by broken lines (fig. 4, p. 141).

(d) *Motion of a rigid body about a fixed point*

The motion of a rigid system about a fixed point bears the same relation to plane motion as spherical geometry bears to plane geometry. Just as the points on a normal to the fixed plane describe congruent orbits in the case of plane motion of a rigid body, so here the points on a radius through the fixed centre describe paths which are similar spherical curves, similarly placed with respect to the fixed point. It is therefore sufficient to examine the motion on any spherical surface having the fixed point as centre. We may thus picture the situation as the sliding of a spherical shell upon an inner fixed sphere. We can apply the same construction as we used for plane motion, and so obtain an instantaneous centre on the sphere. If we trace the path of the instantaneous centre, we obtain two spherical curves, the *herpolhode* and the *polhode*. By rolling the latter curve upon the former, the motion is determined exactly as in plane motion. In reality, we are not dealing with two spherical curves, one rolling on the other, but with two cones, if we consider all the concentric spheres. In the same way, in the case of plane motion, we really have two cylinders, whose generators are normal to the plane, rolling one upon the other.

(e) *Arbitrary motion of a rigid body*

Since the position of a rigid body is determined by the position of one point, and by the orientation of a set of axes fixed in the body, a change to a neighbouring position can be effected by a translatory motion which brings that point to its new position, followed by an infinitesimal rotation of the axes to their new position. We can always arrange matters so that the axis of instantaneous rotation lies in the direction of the translational motion, and a rigid body may therefore always be moved to a neighbouring position by means of an infinitesimal screw motion. We give this result without proof, since no further use will be made of it. If we again trace these instantaneous screw axes in both the moving and fixed systems, we obtain two ruled surfaces. In the course of the motion, these surfaces not only roll upon



one another, as in plane motion or spherical motion, but there is a simultaneous slipping along the instantaneous common generator.

## 2. General Statics and Dynamics of Rigid Bodies. Equivalence of Systems of Forces acting upon Rigid Bodies.

According to Chap. V (p. 85), the necessary and sufficient condition for the equilibrium of a single particle (i.e. for the absence of all accelerations) is the vanishing of the resultant of all applied forces. In the general mechanics of systems of particles, the vanishing of the resultants of all force vectors means that there is no acceleration of the centre of gravity of the system: i.e. if the centre of gravity was at rest *before* the system of forces having zero resultant was applied, it remains at rest when the set of forces is applied. This still does not say that the system of particles remains at rest, for rotational accelerations about the centre of gravity may still take place. For a rigid body, the vanishing of the resultant moment

$$\mathbf{M} = \Sigma [\mathbf{r}_i \mathbf{F}_i]$$

is the necessary and sufficient condition for the absence of this kind of acceleration also. The reason is that the rigid connexions of the separate parts ensure that the changes in angular momenta of the parts are all of the same sign, so that it is impossible for these changes to compensate each other, as may occur in an arbitrary system of particles. As a result, when  $\mathbf{M} = 0$ , equation (7) (p. 110) becomes

$$0 = \frac{d\mathbf{P}}{dt} = \frac{d}{dt} \Sigma \mathbf{p}_i = \frac{d}{dt} \Sigma m_i [\mathbf{r}_i \mathbf{v}_i].$$

As we have just seen, the absence of a change in the total angular momentum  $\mathbf{P}$  implies that there is no change in the angular momentum of the separate parts. If then  $\mathbf{F}$  also vanishes, and no motion existed at first, the system remains at rest. The necessary and sufficient conditions for the equilibrium of a system of forces applied to a rigid body are therefore the vanishing of the resultant force and of the resultant moment of the forces:

$$\Sigma \mathbf{F}_i = \mathbf{F} = 0 \quad \text{and} \quad \Sigma [\mathbf{r}_i \mathbf{F}_i] = \mathbf{M} = 0. \quad . \quad . \quad (5)$$

The effect of a system of forces on a rigid body depends solely upon these two resultants. This is evident for statics, but in kinetics, too, only these two vectors occur, as we shall see in the following sections. All systems of forces which yield the same resultants  $\mathbf{F}$  and  $\mathbf{M}$  are entirely equivalent in effect.

While the vector quantities with which we have dealt thus far could be moved about parallel to themselves at will, without changing their significance, this is no longer true of the forces acting upon a rigid body. It is evident that if we displace parallel to itself a force

vector acting at a point  $P_i$ , we change the point of application in such manner that the moment of the force is altered. The moment is unchanged only if the displacement is *along* the vector. A force acting on a rigid body is a vector localized in a line and may be moved only along its line of action.

The value of the resultant moment depends, in general, upon the position of the reference point  $O$ . There is an important special case where this is not so, viz. if the resultant  $\mathbf{F}$  vanishes. If, for example, we shift the reference point to  $O'$  by a vector  $\mathbf{d}$ , and denote the vectors radiating from  $O'$  by primes, then

$$\begin{aligned}\mathbf{r}' &= \mathbf{r} - \mathbf{d}, \text{ i.e. } \Sigma[\mathbf{r}'_i \mathbf{F}_i] = \Sigma[\mathbf{r}_i \mathbf{F}_i] - \Sigma[\mathbf{d} \mathbf{F}_i] \\ &= \Sigma[\mathbf{r}_i \mathbf{F}_i] - [\mathbf{d} \mathbf{F}]. \quad . . . (6)\end{aligned}$$

The expression  $\Sigma[\mathbf{d} \mathbf{F}_i]$  vanishes, no matter what the direction of  $\mathbf{d}$ , if  $\Sigma \mathbf{F}_i = 0$ , so that the centre of gravity remains at rest or moves uniformly.

The scalar product  $\mathbf{F} \mathbf{M}$  is independent of the point of reference in any case, for if we multiply both members of equation (6) scalarly by  $\mathbf{F}$ , the term  $\mathbf{F}[\mathbf{d} \mathbf{F}]$  vanishes, and there remains

$$\mathbf{F} \Sigma[\mathbf{r}'_i \mathbf{F}_i] = \mathbf{F} \Sigma[\mathbf{r}_i \mathbf{F}_i]. \quad . . . . . (7)$$

It may be shown also that there is always an axis such that if any point thereon be chosen as reference point,  $\mathbf{M}$  will have the direction of  $\mathbf{F}$ . This axis is called the *central axis* of the system of forces. If we denote the moment referred to  $O'$  by  $\mathbf{M}'$ , the vector  $\mathbf{M}'$  should be so determined that

$$\mathbf{M}' = \mathbf{M} - [\mathbf{d} \mathbf{F}] = \gamma \mathbf{F}, \quad . . . . . (8)$$

where  $\gamma$  is a scalar which is yet undetermined. Scalar multiplication by  $\mathbf{F}$  yields

$$\mathbf{M}' \mathbf{F} = \mathbf{M} \mathbf{F} = \gamma \mathbf{F}^2,$$

that is,

$$\gamma = \frac{\mathbf{M} \mathbf{F}}{\mathbf{F}^2}$$

$$\text{and} \quad [\mathbf{d} \mathbf{F}] = \mathbf{M} - \gamma \mathbf{F} = \mathbf{M} - \frac{\mathbf{F}(\mathbf{M} \mathbf{F})}{\mathbf{F}^2} = \frac{\mathbf{M}(\mathbf{F}^2) - \mathbf{F}(\mathbf{M} \mathbf{F})}{\mathbf{F}^2} = \frac{[\mathbf{F}(\mathbf{M} \mathbf{F})]}{\mathbf{F}^2}. \quad (9)$$

From (9) we see that

$$\mathbf{d} = - \frac{[\mathbf{M} \mathbf{F}]}{\mathbf{F}^2} + \lambda \mathbf{F}, \quad . . . . . (10)$$

where  $\lambda$  is an arbitrary variable scalar quantity. This means, however, that the terminus of  $\mathbf{d}$  is on a straight line parallel to  $\mathbf{F}$ . If, in particular,  $\mathbf{M}$  is parallel to  $\mathbf{F}$ , the central axis passes through  $O$  and is parallel to  $\mathbf{F}$ .

The simplest system of forces having a vanishing resultant  $\mathbf{F}$  but a non-vanishing  $\mathbf{M}$  consists of two equal and oppositely directed forces,  $\mathbf{F}$  applied at the point  $\mathbf{r}_1$  and  $-\mathbf{F}$  at  $\mathbf{r}_2$ . This system is called a *couple*. The resultant moment  $\mathbf{M} = [(\mathbf{r}_1 - \mathbf{r}_2) \mathbf{F}]$  is a vector perpendicular to the plane of the two parallel forces and equal in magnitude to  $F$  times the perpendicular distance between the forces. Since we saw



that the effect of a set of forces applied to a rigid body depends only upon  $\mathbf{F}$  and  $\mathbf{M}$ , we see that any system of forces may be replaced by a single force  $\mathbf{F}$  and a couple  $\mathbf{M}$  whose moment is equal to the resultant moment of the set of forces.

If all the forces acting on a rigid body are coplanar—a frequent occurrence in practice—then the magnitude and line of action of their resultant may be found in a vivid way by means of two graphical constructions known as the *polygon of forces* and the *funicular polygon*. In fig. 5a the four applied forces  $\mathbf{F}_1, \mathbf{F}_2, \mathbf{F}_3, \mathbf{F}_4$  are shown each in its correct line of action. In fig. 5b the resultant  $\mathbf{F}$  is determined by the familiar force polygon construction. In order to find the correct line of action of the resultant, we select an arbitrary point  $O$  and draw lines from it to the various corners of the polygon. These lines may be thought of as auxiliary force vectors  $\mathbf{F}_{0,1}, \mathbf{F}_{1,2}, \mathbf{F}_{2,3}, \mathbf{F}_{3,4}$  and  $\mathbf{F}_{4,0}$ . From the triangles thus formed it appears that each pair of auxiliary forces is in equilibrium with one of the given forces, provided that we apply both of the former at a single point on

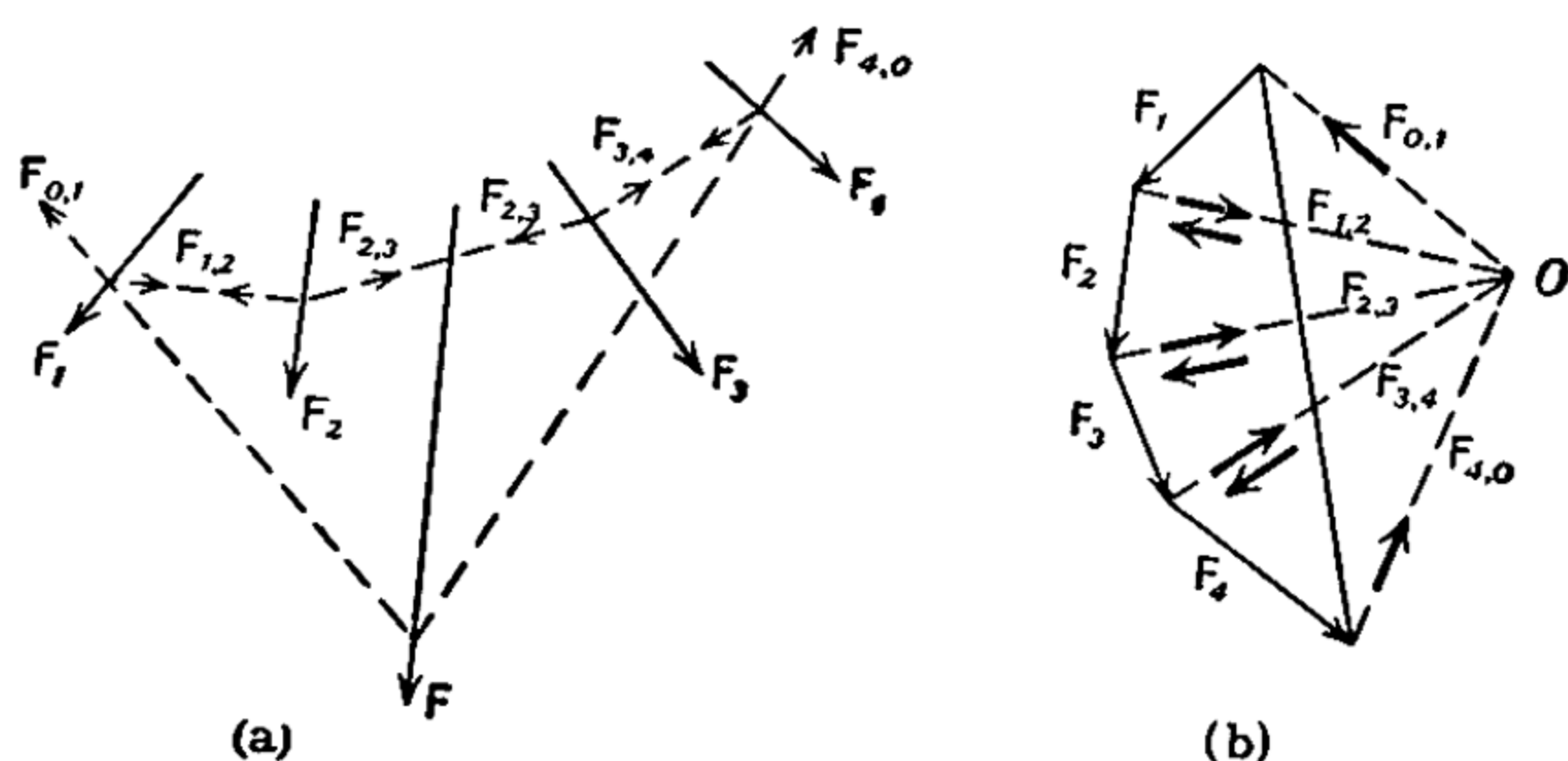


Fig. 5

the line of action of the given force. Except for  $\mathbf{F}_{0,1}$  and  $\mathbf{F}_{4,0}$ , however, all auxiliary forces enter twice, each time with opposite signs; and since the two members of a pair act in the same line they cannot change anything connected with the equilibrium conditions. Inasmuch as equilibrium obtains after the introduction of the auxiliary forces,  $\mathbf{F}_{0,1}$  and  $\mathbf{F}_{4,0}$  must be in equilibrium with the resultant  $\mathbf{F}$ , and the line of action of the latter must pass through the intersection of those of the former two forces. Hence in diagram (a), through any point of  $\mathbf{F}_1$  we draw a line parallel to  $\mathbf{F}_{0,1}$  and another parallel to  $\mathbf{F}_{1,2}$ . Through the intersection of the latter with  $\mathbf{F}_2$  we draw another line parallel to  $\mathbf{F}_{2,3}$ , and so on. The intersection of  $\mathbf{F}_{0,1}$  and  $\mathbf{F}_{4,0}$  is a point on the resultant, whose direction is that of the closing side  $\mathbf{F}$  of the polygon in (b).

This sequence of lines supplied to the diagram of original forces is called the *funicular* (Latin: pertaining to a rope) *polygon* because a rope would assume this shape under the action of the given forces. In the event that the force polygon happens to be self-closing, the directions of  $\mathbf{F}_{0,1}$  and  $\mathbf{F}_{4,0}$  will be parallel and the resultant will be a couple. If the two parallel lines are coincident, in which case the funicular polygon will be closed also, the given forces will be in equilibrium.

**Ex. 47.** A beam is supported at its ends and carries a concentrated load of  $P$  kg. at a point  $\frac{1}{3}$  the distance from one end. By reversing the construction just described, find the magnitude of the downward forces on the supports. How does one proceed when several loads are applied?



### 3. Rotation of a Rigid Body about a Fixed Axis. Moment of Inertia and its Calculation.

We begin the subject of the dynamics of a rigid body with the simplest case—"one-dimensional" motion about a fixed axis. This motion is determined by a single co-ordinate, the angle of turn  $\phi$ . Thinking of the axis as a material *axle*, we may consider it fixed by holding two of its points,  $O$  and  $O'$ , in bearings. Let the resultant of the external forces be  $\mathbf{F}$ , and the resultant moment,  $\mathbf{M}$ . If we apply the general equations of the mechanics of systems of particles, thus treating the body as free, we must also take into account the reactions at the bearings. That is to say, if the bearings were not present, the axis of rotation would, in general, tend to move in some direction. This is prevented by the bearings, and so the axle exerts a pressure on the bearings. We may consider the bearings replaced by forces which are equal to and opposite to the pressure exerted by the axle at any instant. Call these forces, acting at  $O$  and  $O'$ ,  $\mathbf{F}_r$  and  $\mathbf{F}_r'$  respectively. If we put the origin of co-ordinates at  $O$ , the point of application of one of these forces, then  $\mathbf{r}\mathbf{F}_r$  vanishes, and the equation for the rate of change of the total angular momentum of the system becomes

$$\frac{d\mathbf{P}}{dt} = [\mathbf{r}'\mathbf{F}_r'] + \mathbf{M}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (11)$$

We now put the  $z$ -axis of our co-ordinate system along the axis of rotation. Since the vector product  $[\mathbf{r}'\mathbf{F}_r']$  has no  $z$ -component, on account of  $\mathbf{r}' = r'\mathbf{k}$ , the reaction force for the  $z$ -component of equation (11) drops out, and we obtain the simpler equation

$$\frac{dP_z}{dt} = M_z, \quad . \quad . \quad . \quad . \quad . \quad . \quad (12)$$

which we will now treat further. If the angular velocity vector is

$$\boldsymbol{\omega} = \omega\mathbf{k}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (13)$$

then

$$\begin{aligned} \mathbf{P} &= \sum m_i [\mathbf{r}_i \mathbf{v}_i] = \sum m_i [\mathbf{r}_i [\boldsymbol{\omega} \mathbf{r}_i]] \\ &= \boldsymbol{\omega} \sum m_i r_i^2 - \sum m_i \mathbf{r}_i (\boldsymbol{\omega} \mathbf{r}_i) \\ &= \omega \mathbf{k} \sum m_i r_i^2 - \omega \sum m_i \mathbf{r}_i (\mathbf{r}_i \mathbf{k}), \quad . \quad . \quad . \quad (14) \end{aligned}$$

and

$$\begin{aligned} P_z &= \omega \sum m_i r_i^2 - \omega \sum m_i (\mathbf{r}_i \mathbf{k})^2 = \omega \sum m_i [\mathbf{r}_i \mathbf{k}]^2 \\ &= \omega \sum m_i (x_i^2 + y_i^2). \quad . \quad . \quad . \quad . \quad . \quad . \quad (15) \end{aligned}$$

The quantity

$$\Sigma m_i [\mathbf{r}_i \mathbf{k}]^2 = \Sigma m_i (x_i^2 + y_i^2) = \Sigma m_i R_i^2 \quad . \quad . \quad (16)$$

is called the moment of inertia of the body about the  $z$ -axis, and is denoted by  $I_{zz}$ , for a reason explained below. With this notation, (12) gives the equation of motion

$$I_{zz} \frac{d\omega}{dt} = I_{zz} \frac{d^2\phi}{dt^2} = M_s \quad . \quad . \quad . \quad . \quad (17)$$

Since the position of the rotating body is determined by the angle  $\phi$ , no other equations are needed if we wish to study only the course of the motion. We must bring in the remaining equations of motion, however, if we wish to find the pressures at the bearings. Comparing (17) with the statement of Newton's Second Law for the case of one degree of freedom of a particle, viz.

$$m \frac{d^2x}{dt^2} = F_x,$$

we see that the two equations are mathematically identical, and that the following paired quantities correspond to each other:

Moment of inertia  $I_{zz}$ .

Mass  $m$ .

Angular acceleration  $\frac{d^2\phi}{dt^2}$ .

Linear acceleration  $\frac{d^2x}{dt^2}$ .

Component, along the axis of rotation, of the moment of the external forces.

Component of force in the direction of motion.

Moreover, the kinetic energy of rotation becomes, in complete analogy with the case of linear motion,

$$T = \frac{1}{2} I_{zz} \left( \frac{d\phi}{dt} \right)^2, \quad . \quad . \quad . \quad . \quad (18)$$

as may be verified readily.

In the mechanics of a particle, the vibrational motion taking place under the influence of a quasi-elastic force  $F_x = -kx$  is of importance. In the rotational motion of rigid bodies, also, there often arise restoring torques which are proportional to the angle of turn, and whose representative vectors are in the direction of the axis. Such a turning moment is of the form  $M = -\tau\phi$ , and the equation of motion becomes

$$I \frac{d^2\phi}{dt^2} + \tau\phi = 0, \quad . \quad . \quad . \quad . \quad (19)$$

where in place of  $I_{zz}$  we write  $I$  for brevity, it being understood that we mean the moment of inertia about the axis of rotation. The factor of proportionality  $\tau$  is called the *moment of torsion*; it has the dimensions

of force  $\times$  distance, i.e. work. A torque of this kind arises, for example, when a rigid body, suspended by a wire clamped at the upper end, is slightly twisted about the wire as axis. Since equation (19) is identical with equation (24) (p. 93), apart from notation, we can write down at once the frequency of free vibration of a rigid body under the influence of a restoring force proportional to the angle of twist:

$$\nu = \frac{1}{2\pi} \sqrt{\frac{\tau}{I}} = \frac{1}{2\pi} \sqrt{\frac{\text{moment of torsion}}{\text{moment of inertia}}}. \quad \cdot \cdot \cdot (20)$$

In order to determine  $I$ , we must calculate the quantity  $\Sigma m_i h_i^2$ , where  $h_i$  is the perpendicular distance of the  $i$ th particle from the axis. For a continuous distribution of matter of density  $\rho$ , the sum is to be replaced by the integral  $\iiint \rho h^2 dx dy dz$ . We shall find, below, that the moment of inertia about any axis may be found without another integration, provided that this quantity is known for a particular set of three mutually perpendicular axes of the body intersecting at the centre of gravity. We show in the first place that the moment of inertia about any axis not passing through the centre of gravity may be calculated from the value of the moment of inertia about a parallel axis through the centre of gravity. If  $\bar{\mathbf{r}}$  is the vector to the centre of gravity from an origin on any selected axis, and if we denote by primes the radius vectors from the centre of gravity to the various particles, we have

$$\mathbf{r}_i = \bar{\mathbf{r}} + \mathbf{r}'_i. \quad \cdot \cdot \cdot \cdot \cdot \cdot (21)$$

If the axial direction is denoted by the unit vector  $\mathbf{c}$ ,

$$I = \Sigma m_i [\mathbf{r}_i \mathbf{c}]^2 = \Sigma m_i [\bar{\mathbf{r}} \mathbf{c}]^2 + \Sigma m_i [\mathbf{r}'_i \mathbf{c}]^2, \quad \cdot (22)$$

since  $\Sigma m_i \mathbf{r}'_i = 0$  (cf. p. 107), i.e. the term

$$2\Sigma m_i [\bar{\mathbf{r}} \mathbf{c}] \cdot [\mathbf{r}'_i \mathbf{c}] = 2[\bar{\mathbf{r}} \mathbf{c}] \cdot [\Sigma m_i \mathbf{r}'_i, \mathbf{c}]$$

vanishes.

The second term of (22) represents the moment of inertia referred to the axis passing through the centre of gravity; the first term represents the moment of inertia of the body about the new axis, considering the entire mass concentrated at the centre of gravity. This is *Steiner's Theorem*:

*The moment of inertia of a rigid body referred to an arbitrary axis is equal to the moment of inertia referred to a parallel axis through the centre of gravity, plus the moment of inertia of the entire mass about the given axis, assuming it to be concentrated at the centre of gravity.\**

\* It is sometimes convenient to introduce the so-called *radius of gyration*  $k$  defined by

$$k^2 = \frac{\Sigma m_i h_i^2}{\Sigma m_i} = \frac{I}{\Sigma m_i};$$

i.e.  $k$  is the distance from the axis at which the entire mass may be considered con-





This equation, being of the second degree, represents a conicoid, or quadric surface. It is an ellipsoid (in special cases, an ellipsoid of rotation, or a sphere), since the moment of inertia of a physical body, referred to any axis, cannot vanish, and hence no radius vector  $R = 1/\sqrt{I}$  can become infinite. Since the quantities  $I_{xx}$ , &c., depend on the choice of reference point, there will be a different ellipsoid of inertia for every point of reference  $O$ . Now, according to analytical geometry, for every surface of the second degree there is a particular system of co-ordinates, that of the principal axes, in which the equation of the surface has the simple form

$$I_1 x'^2 + I_2 y'^2 + I_3 z'^2 = 1. \quad . \quad . \quad . \quad (29)$$

The three quantities  $I_1, I_2, I_3$  are called the *principal moments of inertia*. If the axes of the principal moments and the magnitudes are known, the moment of inertia for any other axis through  $O$  may be evaluated graphically or analytically as the inverse square of the radius vector drawn in the direction of the axis in question. If  $O$  is the centre of gravity, the moment of inertia for any other axis may be found by applying Steiner's theorem. A second-degree surface has six independent coefficients. In general, then, one must measure six moments of inertia, in so far as the positions of the principal axes of inertia (which correspond to the axes of symmetry) cannot be determined in advance on grounds of symmetry. The six quantities determining the ellipsoid of inertia are to be interpreted as follows—three for the principal moments of inertia and three additional ones for the orientation of the system of principal axes within the body.

The expression for the kinetic energy of a rotating rigid body is specially simple in the co-ordinate system of the principal axes of inertia. In general we have

$$T = \frac{1}{2} \sum m_i v_i^2 = \frac{1}{2} \sum m_i [\omega r_i]^2,$$

and since

$$\omega = \omega c,$$

this expression becomes

$$T = \frac{1}{2} \omega^2 \sum m_i [c r_i]^2 = \frac{1}{2} \omega^2 I.$$

But we have already calculated the expression for  $\sum m_i [c r_i]^2$  in the co-ordinate system of the principal axes. Since  $I_{kl} = 0$  for  $k \neq l$ , we have from equation (23),

$$T = \frac{1}{2} \omega^2 (I_1 \cos^2 \alpha + I_2 \cos^2 \beta + I_3 \cos^2 \gamma), \quad . \quad . \quad (30)$$

and since

$$\cos \alpha = \frac{\omega_x}{\omega}, \quad \cos \beta = \frac{\omega_y}{\omega}, \quad \cos \gamma = \frac{\omega_z}{\omega}, \quad . \quad . \quad (31)$$

resolution of  $\omega$  along the principal axes of inertia fixed in the body makes the expression for the kinetic energy take the form

$$T = \frac{1}{2}(I_1\omega_x^2 + I_2\omega_y^2 + I_3\omega_z^2). \quad . \quad . \quad . \quad (32)$$

*Ex. 48.* Calculate the moment of inertia of a solid, homogeneous cube about one of its diagonals.

*Ex. 49. Theory of Kater's Pendulum.* If a rigid pendulum has the same period when oscillating about two parallel axes at unequal distances  $s$  and  $s'$  from the centre of gravity, then the length of the simple pendulum having the same period is given by  $s + s'$ . Prove this.

*Ex. 50.* What must be the distance of the point of suspension of a pendulum from its centre of gravity in order that a small change in this distance may have a minimum effect on the period?

*Ex. 51.* For many purposes, a diatomic molecule may be represented by a rigid "dumb-bell" model, in which the masses  $M_1$  and  $M_2$  are a fixed distance  $a$  apart. This system rotates as it flies through space. Represent the kinetic energy as the sum of translational and rotational energy.

#### 4. Motion of a Rigid Body about a Fixed Point. Elements of the Theory of the Top.

##### (a) Relation between angular momentum and angular velocity

For the investigation of the motion of a rigid body about a fixed axis we needed only the component of total angular momentum in the direction of the axis, and not the forces holding the axis in place, so long as we were interested only in the progress of the motion, and not in the forces at the bearings in which the imaginary axle was considered to turn. If, now, one of the bearings is removed, and the axle is constrained only at one point  $O$ , the axis will not, in general, remain stationary, but will turn about  $O$ . If we wish to apply the theorem concerning the total moment of momentum, we must consider again the relation between the angular momentum vector  $\mathbf{P}$  and the angular velocity vector  $\omega$ , given (p. 146) by

$$\mathbf{P} = \sum m_i \mathbf{r}_i^2(\omega) - \sum m_i \mathbf{r}_i(\mathbf{r}_i \omega).$$

Writing the equation in components,

$$P_x = \sum m_i(x_i^2 + y_i^2 + z_i^2)\omega_x - \sum m_i x_i(x_i\omega_x + y_i\omega_y + z_i\omega_z),$$

$$\begin{aligned} \text{or} \quad & P_x = I_{xx}\omega_x - I_{xy}\omega_y - I_{xz}\omega_z \\ \text{and similarly,} \quad & \left. \begin{aligned} P_y &= -I_{xy}\omega_x + I_{yy}\omega_y - I_{yz}\omega_z \\ P_z &= -I_{xz}\omega_x - I_{yz}\omega_y + I_{zz}\omega_z \end{aligned} \right\} . \quad . \quad . \quad (33) \end{aligned}$$

That is, the components of  $\mathbf{P}$  are linear, homogeneous functions of the components of  $\omega$ . The matrix of coefficients is symmetric, since the terms symmetric with respect to the diagonal are equal. Starting with a given vector  $\omega$ , the correspondence between the vectors  $\mathbf{P}$  and  $\omega$ , determined by six coefficients, must, of course, always lead to the same



vector  $\mathbf{P}$ , in whatever way the vectors are resolved into components. But, according to p. 36, this is the criterion that the six coefficients should be the components of a symmetric tensor—the *inertial tensor*. According to the explanation on pp. 36 *et seq.*, one can find the vector  $\mathbf{P}$  corresponding to a given vector  $\boldsymbol{\omega}$  by using the tensor ellipsoid:

$$I_{xx}x^2 + I_{yy}y^2 + I_{zz}z^2 - 2I_{xy}xy - 2I_{yz}yz - 2I_{zx}zx = 1. \quad (34)$$

This ellipsoid is identical with the ellipsoid of inertia discussed in the preceding section. To find  $\mathbf{P}$ , draw a radius vector  $\boldsymbol{r}$  parallel to  $\boldsymbol{\omega}$ . Then  $\mathbf{P}$  has the direction of the normal to the ellipsoid at the terminus of  $\boldsymbol{r}$ . The magnitude of  $\mathbf{P}$  is determined as on p. 37 by the equation upon which the construction of the ellipsoid is based. The normal and the radius vector have, in general, different directions, so that the vector of the resultant angular momentum does not, in general, have the same direction as the axis of rotation. The two are parallel only for the directions of the three principal axes of inertia. If we denote by accents the components referred to this co-ordinate system, which is determined by the form of the body and the distribution of mass, then

$$\left. \begin{aligned} P_{x'} &= I_1 \omega_{x'} \\ P_{y'} &= I_2 \omega_{y'} \\ P_{z'} &= I_3 \omega_{z'} \end{aligned} \right\} \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad (35)$$

This co-ordinate system—in which the axes of co-ordinates are the axes of the ellipsoid of inertia—is fixed in the body, and thus changes direction when the body turns. On the other hand, all the theorems already proved, such as the law of conservation of angular momentum in the absence of external torques, are based on a system of co-ordinates fixed in space. For example, if in this case of  $\mathbf{P}$  fixed in space, we imagine the vector  $\mathbf{P}$  to be watched by an observer on the rotating body, he will observe an apparent motion of this vector.

In mechanics, any rapidly rotating rigid body is called a *top*. The easiest case to treat is that of a top having three equal moments of inertia. It is called a *spherical top*. Naturally, the body itself need not be a sphere; it is only necessary that the mass be so distributed as to make the ellipsoid of inertia a sphere. By (35),  $\mathbf{P}$  and  $\boldsymbol{\omega}$  will have the same direction for a spherical top. For other reasons, the tops or *gyroscopes* encountered in practice are rotationally symmetric bodies having two of their principal moments of inertia equal, with the consequence that the ellipsoid of inertia is an ellipsoid of rotation. Tops having three different moments are very difficult to handle mathematically and almost never occur in engineering practice. They have their place, nevertheless, as models for unsymmetric polyatomic molecules. We shall restrict our attention to the symmetric top.

Let us choose the  $k'$ -axis as the axis of symmetry of the rotor,

also called the *axis of figure*. This will also be the axis of figure of the inertial ellipsoid. Call  $I$  the moment of inertia *about* the figure axis, and  $J$  that about an axis *normal* to the figure axis. Call the ratio  $I/J = \theta$ . The numerical value of  $\theta$  must be between 0 and 2, as shown by the following reasoning: on account of the rotational symmetry,  $x' = y'$ , and we have

$$I = \int (x'^2 + y'^2) dm = 2 \int x'^2 dm, \quad . \quad . \quad . \quad (36)$$

$$J = I/\theta = \int (x'^2 + z'^2) dm = \frac{1}{2}I + \int z'^2 dm. \quad . \quad . \quad (37)$$

Since  $I/2$  occurs in (37) with a positive quantity added to it, we must have  $1/2 < 1/\theta < \infty$ , or  $2 > \theta > 0$ .

For a symmetrical top,  $k'$ , the moment of momentum  $\mathbf{P}$ , and the instantaneous angular velocity vector  $\boldsymbol{\omega}$  all lie in one plane; for the construction of  $\mathbf{P}$  by means of  $\boldsymbol{\omega}$ , using the inertial ellipsoid, is carried out in a meridian plane. The same result follows analytically provided that  $\boldsymbol{\omega}$  can be written as the sum of a  $\mathbf{P}$  component and a  $k'$  component. First, if we write  $\boldsymbol{\omega}$  and  $\mathbf{P}$  in component form, referred to the set of axes fixed in the body, and if we combine the components of  $\boldsymbol{\omega}$  that are normal to  $k'$  to form a vector  $\boldsymbol{\omega}_n$ , we obtain

$$\boldsymbol{\omega} = \omega_x i' + \omega_y j' + \omega_z k' = \boldsymbol{\omega}_n + \omega_z k', \quad . \quad . \quad (38)$$

$$\mathbf{P} = \frac{I}{\theta} (\omega_x i' + \omega_y j') + I \omega_z k' = \frac{I}{\theta} \boldsymbol{\omega}_n + I \omega_z k'. \quad (39)$$

Eliminating  $\boldsymbol{\omega}_n$ ,

$$\boldsymbol{\omega} = \frac{\theta}{I} \mathbf{P} - (\theta - 1) \omega_z k'. \quad . \quad . \quad . \quad (40)$$

Fig. 6 represents graphically the content of equations (38) to (40).

### (b) Symmetrical top subject to no forces

Consider now the behaviour of a symmetrical top on which no force moment acts. In this instance the vector  $\mathbf{P}$  will be constant, and so differentiation of (40) gives

$$\frac{d\boldsymbol{\omega}}{dt} = (1 - \theta) \frac{d\omega_z}{dt} k' + (1 - \theta) \omega_z \frac{dk'}{dt}.$$

But according to p. 140 the rate of change  $d\mathbf{r}/dt$  of a radius vector  $\mathbf{r}$  due to rotation is given by  $[\boldsymbol{\omega}\mathbf{r}]$ ; and since any vector, including the axial unit vector  $k'$ , may be represented as the difference of two radius vectors, we have

$$\frac{dk'}{dt} = [\boldsymbol{\omega}k'] \quad \text{and} \quad \boldsymbol{\omega} \frac{dk'}{dt} = 0. \quad . \quad . \quad . \quad (41)$$

Hence 
$$\frac{d\omega}{dt} = (1 - \theta) \frac{d\omega_z}{dt} k' + (1 - \theta)\omega_z [\omega k']. \quad (42)$$

It follows from the last relation that  $\omega_z$  is constant; for if we form the scalar product of this equation with  $k'$ , it follows from (41) that with  $k'\omega = \omega_z$ , that  $d\omega_z/dt = 0$ . However, not only  $\omega_z$  but also the magnitude  $\omega$  of  $\omega$  is constant; for if we form the scalar product of (42) with  $\omega$ , and use the fact that  $\omega_z = \text{const.}$  because  $\omega[\omega k'] = 0$ , we get

$$\omega \frac{d\omega}{dt} = \frac{d(\omega^2/2)}{dt} = 0. \quad (43)$$

But since  $\omega_z = \omega k'$ , the angle between  $\omega$  and  $k'$  must also be constant. According to fig. 6, which is valid at any instant, the angle

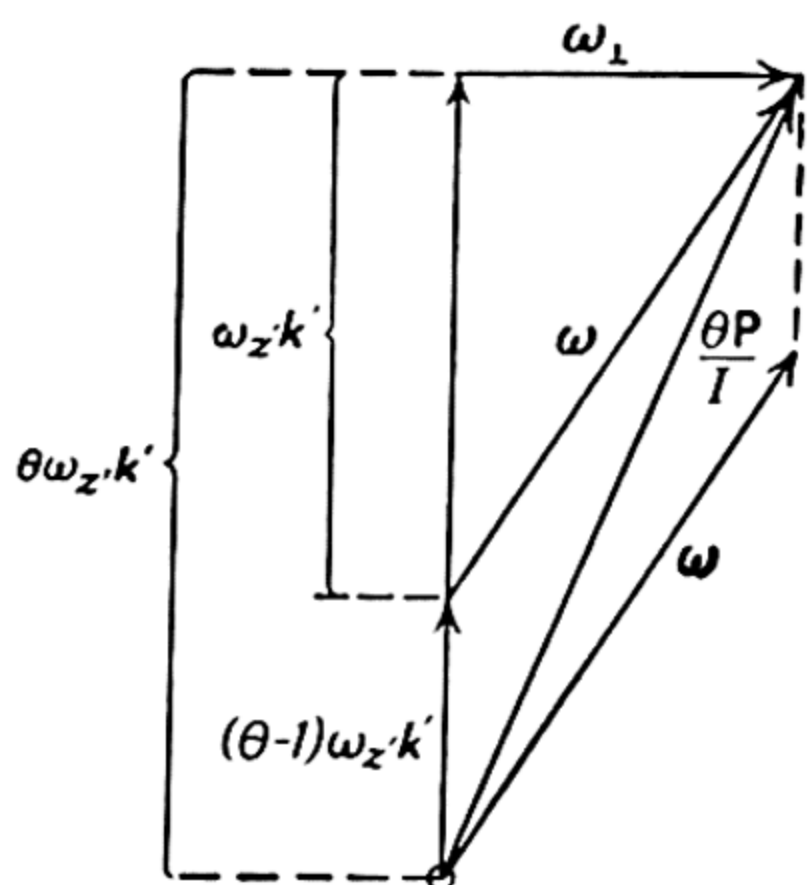


Fig. 6

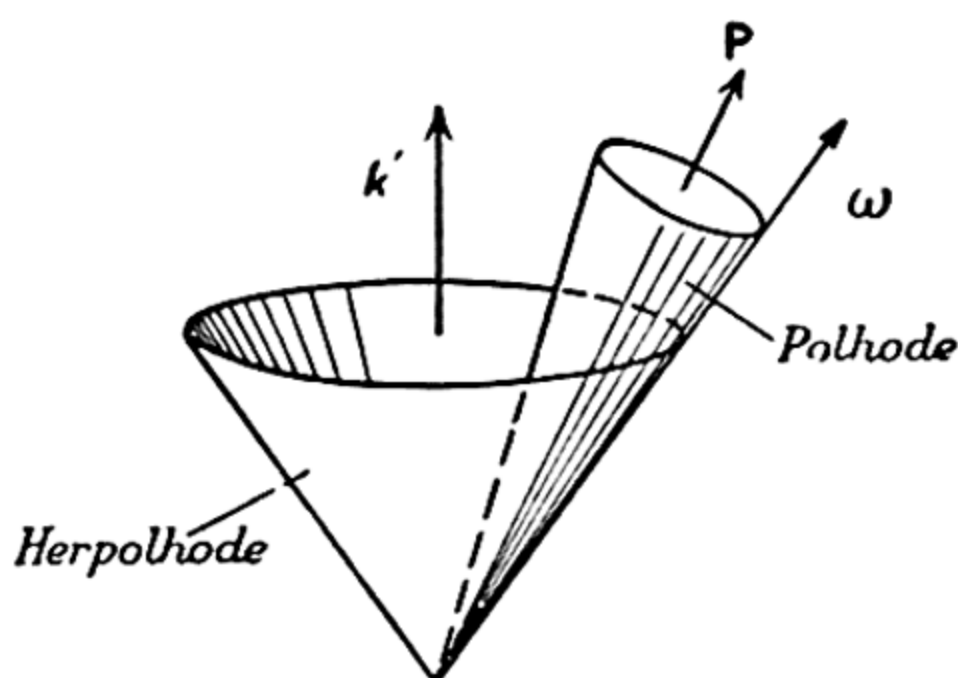


Fig. 7

formed by  $P$  and  $k'$  is also constant, and the only possibility of a change in the vectors is for the plane of the figure to rotate. The axis for such a rotation must be the invariable line  $P$ . The general motion of a force-free symmetrical top is called *nutation*. In it, the instantaneous axis  $\omega$  describes a cone about the axis  $P$ , which is fixed in space, while the top itself rotates about  $\omega$ . On the other hand, any motion of a rigid body about a fixed point may be represented according to § 1 (c) by the rolling of the polhode on the herpolhode, the instantaneous angular velocity being that of the generator common to both cones. Thus in the present instance the herpolhode is the cone described about the axis  $P$  by  $\omega$ , while the polhode is described by the motion of  $\omega$  around  $k'$  (fig. 7). Both are right circular cones. Since an actual gyroscope always has high speed about the figure axis, and since the nutation arises from disturbances, as will be explained below, the angles between  $k'$ ,  $P$  and  $\omega$  are small because  $P$  has very nearly the same direction as  $k'$ . However, the angular speed of rotation of  $k'$  around the fixed axis  $P$  is comparable with  $\omega_z$ , and so is large. This



follows from the fact that the component in question is, by (40),  $(\theta/I)\mathbf{P}$ . Taking as an approximation  $|\mathbf{P}| = I\omega_z$ , the magnitude of this component becomes  $\theta\omega_z$ . In a practical case,  $\theta$  does not differ much from unity.

*Example: The wandering of the earth's pole.* The earth is a top whose angular speed is small but whose moment of momentum is large, due to its tremendous rotation inertia. The polar flattening makes  $\theta = 301/300$ . Displacements of matter cause slight irregular wandering of the position of the pole over the earth's surface within a distance of about four metres! According to the preceding development, the angular speed around  $\mathbf{P}$  will amount to  $2\pi(301)/300 \text{ day}^{-1}$ . This is the speed with which an observer fixed in space would see the axis of the earth rotating about  $\mathbf{P}$ . Referring to a terrestrial observer, we must subtract the normal daily rotation rate of  $2\pi \text{ day}^{-1}$ , leaving  $2\pi/300 \text{ day}^{-1}$ , or a period of 300 days. The observed period is 420 days, the departure being due to the deformation of the rotating earth.

Let us examine one more special case of the motion of a force-free top. Suppose  $\omega$  to be constant in *direction* as well as in magnitude. In this case equation (42) can be satisfied only if  $\omega$  and  $\mathbf{k}'$  are parallel, in which case (40) shows that  $\mathbf{P}$  must be parallel to  $\mathbf{k}'$  also. Thus, if the axis of rotation coincides with the axis of figure, the top can continue to rotate about this axis indefinitely. This is the origin of the term "moments of deviation" used in German works to denote the products of inertia; for any axis other than the figure axis these mixed terms do not vanish, and rotation about such an axis is accompanied by "wandering" of the axis of figure.

### (c) *Symmetrical top under the influence of external moments*

Imagine the top  $K$  to be contained in a housing  $G$ , fig. 8, which itself is free to move in all directions—for instance, by being allowed to float in a liquid. The centre of gravity of the top itself is assumed to coincide with the centre of the indicated sphere. Let the axis of the top be again the  $\mathbf{k}'$  axis. First of all, the top must be set into rotation by a force moment having the direction of  $\mathbf{k}'$ . Once the desired rate of rotation is attained, this moment may be gradually decreased. Then, if friction is negligible, the top will continue to run indefinitely at this speed. Thus it is assumed that any moments thereafter applied to the top shall not have  $\mathbf{k}'$  components. Under such conditions  $\omega_z$  will be constant whether or not external moments exist. To see this, set the value of  $\mathbf{P}$  from (40) into the fundamental equation (7), p. 110, obtaining

$$\mathbf{M} = \frac{d\mathbf{P}}{dt} = \frac{I}{\theta} \frac{d\omega}{dt} + \frac{I}{\theta} (\theta - 1) \frac{d\omega_z}{dt} \mathbf{k}' + \frac{I}{\theta} (\theta - 1) \omega_z \frac{d\mathbf{k}'}{dt}. \quad (44)$$

Scalar multiplication by  $\mathbf{k}'$  yields

$$M_z = 0 = \frac{I}{\theta} \frac{d\omega_z}{dt} + \frac{I}{\theta} (\theta - 1) \frac{d\omega_z}{dt}, \quad \text{or} \quad \frac{d\omega_z}{dt} = 0, \quad . \quad (45)$$



we obtain these important forms for the fundamental equation

$$\mathbf{M} = \frac{I}{\theta} \frac{d\Omega_n}{dt} + I\omega_z \frac{d\mathbf{k}'}{dt} = \frac{I}{\theta} \frac{d\Omega_n}{dt} + I\omega_z [\Omega_n \mathbf{k}']. \quad (48)$$

For any practical gyroscope,  $I\omega_z$  represents the main part of the angular momentum; let this quantity be denoted by  $\bar{P}$ .

Consider the following special cases:

(1) *Uniform rotation of the housing about a fixed axis.*

This type of motion of the housing, and hence of the axis of the top, is much more important than the nutation described above. It is called *precession*, and is the most obvious of all gyroscopic phenomena. Assume a co-ordinate system  $i'', j'', k''$  connected with the housing, and let the  $k''$ -axis coincide with the axis  $k'$  of the rotor (fig. 8). Since the component of  $\Omega$  parallel to  $k'$  does not contribute to the vector product  $[\Omega k']$ , equations (48) become for this case

$$\mathbf{M} = \bar{P} \frac{d\mathbf{k}'}{dt} = \bar{P} [\Omega \mathbf{k}']. \quad (48')$$

The change  $d\mathbf{k}'/dt$  in the unit vector  $k'$  must be normal to  $k'$ . Further,  $k'$  rotates as time passes, so that the applied moment  $\mathbf{M}$  must have this same rotation. This will be true, in the case of a force having a constant direction in space, if its line of action does not pass exactly through the centre of gravity of the top. For example, as in fig. 8, let a mass  $m$  be suspended from the housing at one end of the  $j''$ -axis, the axis of the rotor being in an inclined position. Denoting the unit vector of gravity by  $g_0$  and writing  $\delta$  for the angle between  $g_0$  and  $j''$  (which is the same as the angle of inclination of the rotor axis), we have

$$\mathbf{M} = mga [j'' g_0] = -mga (\sin \delta) i''; \quad (49)$$

i.e. the moment always remains perpendicular to  $k''$  (or to  $k'$ ).

For a given moment, equation (48') in itself yields only the magnitude of the component of the angular speed that is normal to  $k'$ . The direction of the axis about which the housing continuously turns is obtained as follows:

Since  $d\mathbf{k}'/dt$  lies in the direction of  $\mathbf{M}$ , then according to (48') and (49),  $d\mathbf{k}'/dt$  must be normal to  $g_0$ ; hence  $(d\mathbf{k}'/dt) g_0 = 0$ . Integration yields

$$k' g_0 + \cos(\pi/2 - \delta) = \text{const.} \quad (50)$$

Thus the axis of the rotor must turn in such a way as to maintain a constant angle with the direction of gravity, i.e. the axis must de-



scribe a cone about the vertical. The magnitude of the precessional angular velocity is, by (48') and (49),

$$\Omega = \frac{\Omega_n}{\cos \delta} = \frac{M/\bar{P}}{\cos \delta} = \frac{mga \tan \delta}{\bar{P}}. \quad \cdot \cdot \cdot \quad (51)$$

A gyroscope, then, does not yield to gravity as does a non-rotating body—it “turns aside at a right angle”. While section (b) showed that the aperture of the *cone of nutation* remains small for small disturbances, the aperture of the *cone of precession* is determined *a priori*. Even when the applied moment is small, an easily perceived change in direction of the axis soon results. It is only the speed of precession that becomes small as the moment is reduced.

## (2) *Nutation of a force-free top produced by impulses.*

Let a sudden moment  $\mathbf{M}$  be applied at time  $t = 0$  to an otherwise force-free top rotating about its axis of figure, and let this moment continue to act for  $\Delta t$  seconds. Denoting the time interval of the moment by  $\mathbf{N}$ , integration of equation (48) with respect to the time gives

$$\mathbf{N} = \int_0^{\Delta t} \mathbf{M} dt = \frac{I}{\theta} \Omega_{n'} + I\omega_z \int_0^{\Delta t} \frac{d\mathbf{k}'}{dt} dt = \frac{I}{\theta} \Omega_{n'} + I\omega_z \Delta \mathbf{k}'. \quad (52)$$

For any impulse, the change in *velocity* in a short time  $\Delta t$  is considerable, while the change in *position* is small to a higher order. For this reason,  $\Delta \mathbf{k}'$  may be neglected. Hence, after impact, a new component of rotation normal to the figure axis and of amount  $\Omega_{n'} = \mathbf{N}\theta/I$  is present. This gives rise to the case of nutation of a force-free top discussed in (1), the quantity  $\Omega_s$  here corresponding to  $\omega_n$  there. Apart from external impacts, internal rearrangements of matter can also act to deflect the axis of  $\mathbf{P}$  from the  $\omega$  direction [cf. (b) above].

Precession and nutation may exist at the same time, as when a precessing top is subjected to external impulses, or when shifting of part of the mass takes place. An impulse also must be considered to operate at the instant when the moment responsible for precession is suddenly applied, for a short-lived force may be represented by sudden application of a steady force, followed immediately by the application of the reaction force. Since the cone of nutation always has a small aperture, the nutation has only the effect of producing a corrugation of the surface of the cone of precession. This is usually negligible in technical applications.

## (3) *Example of the precession of the axis of the earth.*

As a result of the polar flattening of the earth and the inclination of its axis (“obliquity of the ecliptic”) the attraction of the sun gives rise to a turning

moment when the earth is near either end of its orbit (fig. 9a). This moment, which seeks to bring the equator closer to the plane of the orbit, vanishes at the equinoctial positions. However, since the angular momentum of the earth is so large, we know that the precession will be very slow and an average moment may be assumed in computing the actual precession. To be added to this is the average moment due to the moon's attraction. Because of the smaller distance, the latter moment is of the same order of magnitude as that of the sun. The averaging is best done by assuming the masses of sun and moon distributed uniformly along the orbits they appear to have as seen from the earth. For this purpose it is sufficient to assume circular orbits and to take their planes to be coincident. As seen from fig. 9b, the gravitational moment seeks to bring the earth's axis into coincidence with the normal to the orbit plane, so that the vector represent-

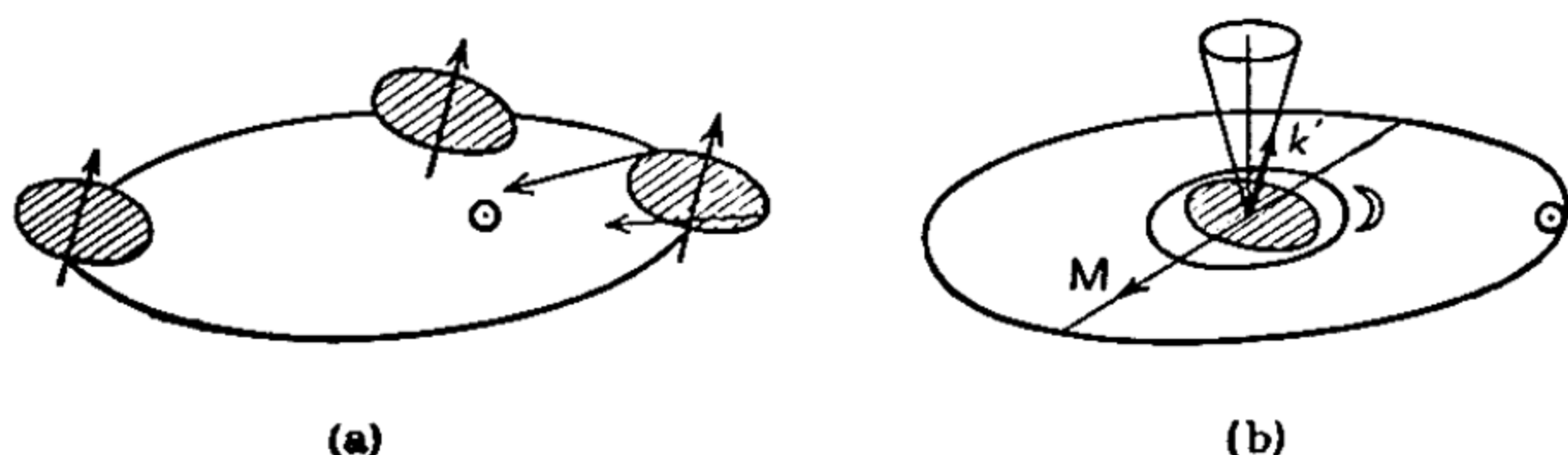


Fig. 9

ing this moment must be perpendicular to the plane containing the earth's axis and the normal. Hence this vector turns in the orbit plane with the precessional motion, just as we saw above that the gravity moment vector acting on the gyroscope turned in the horizontal plane. As in that instance, the earth's axis describes a cone about the normal to the orbit plane. The numerical value of the period of precession turns out to be about 27,000 years.

Due to the fluctuating value of the moment, there is a resulting astronomical nutation. This has no connexion with the wandering of the pole, which was explained above in terms of internal shifting of terrestrial matter.

#### (4) *Example of the gyrocompass.*

Suppose a gyroscope to be mounted in a frame capable of turning only about a vertical axis  $\alpha$ , fig. 10. If the earth did not rotate, the gyroscope would maintain the direction of its axis of rotation, for then it would not be subject to any force moment. On account of the earth's rotation around  $k$  with angular speed  $\omega$ , the mounting—assumed for the present to be fixed to the earth—will be turned; and according to the third law, the rotor will exert a moment  $M'$  on the mounting. This moment is opposite to that which the turning of the axis would produce, and so by (48') it will amount to

$$M' = -\bar{P}\omega[kk']. \quad \dots \dots \dots (53)$$

For an arbitrary position of the axis in a horizontal plane, this moment will have a vertical component  $M_v'$  as well as a horizontal one,  $M_h'$ . The former, according to § 3, produces rotation about  $\alpha$ , while the latter will simply be borne by the bearing  $L$ , because rotation about a horizontal axis is not possible. However, if the axis of the top lies in the meridian plane determined by the vertical and the earth's axis (position  $k'$  of the latter), then vector  $M_h'$  will be horizontal. Since there is no vertical component, the frame will not turn about  $\alpha$ . In other words, the north-south position is one of equilibrium. The stable one of these two direc-

tions is that having the smaller angle between  $k$  and  $k'$ . Because of the smallness of  $\omega$ , the acting moment is so small that it is able to overcome friction in the bearings only with great difficulty. The actual set-up resembles fig. 8. To compensate the moment exerted by the rotor on the housing, one need only attach to the latter a weight  $m$  at a point in the plane normal to the rotor axis. In order

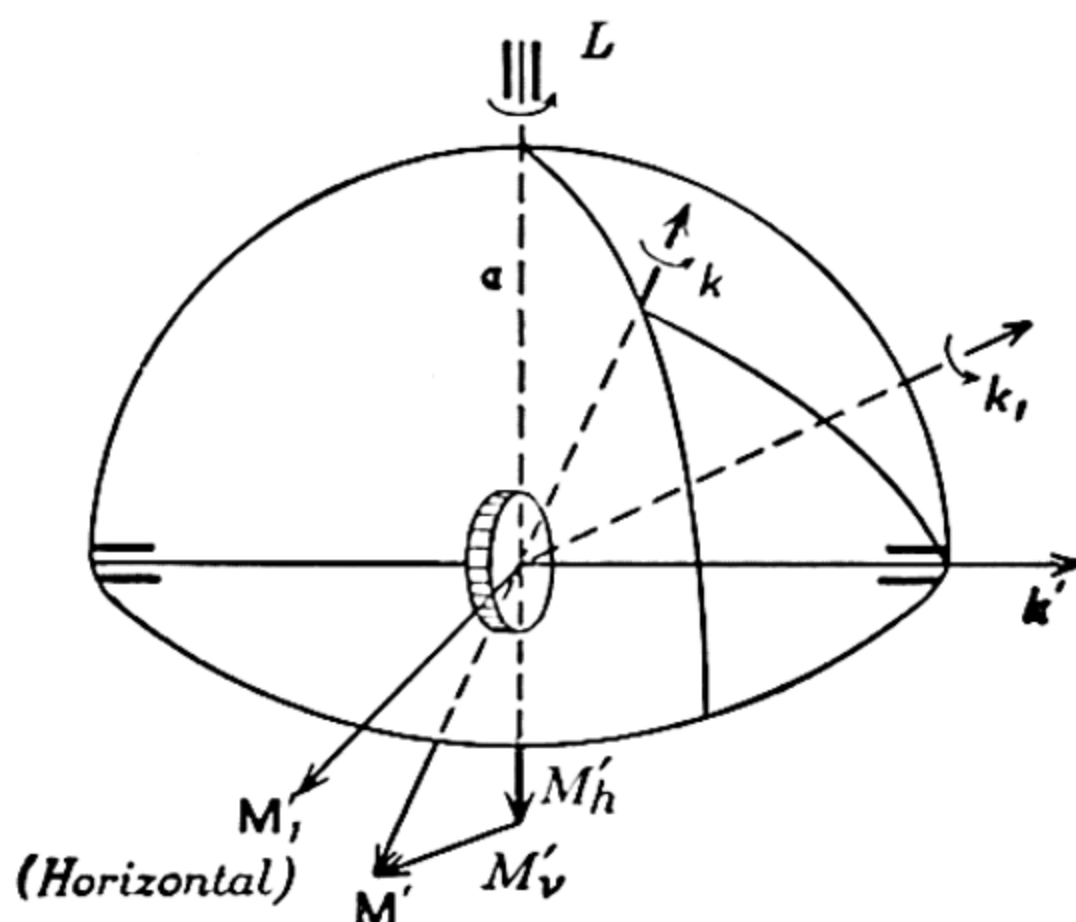


Fig. 10

that this weight may exert a moment, the rotor axis must not be exactly horizontal, but must make a small angle  $\delta$  with this direction. Then, using the notation of fig. 10,

$$mga[j''g_0] + \mathbf{M}' = mga[j''g_0] - \bar{P}\omega[kk'] = 0. \quad . \quad . \quad (54)$$

Since a product vector is normal to the plane of its two factors, equality of the products signifies that the vectors  $k$ ,  $k'$ ,  $j''$  and  $g_0$  are coplanar, i.e. the rotor axis must again lie in the meridian plane when equilibrium is attained.

*Ex. 52.* In the arrangement just described, compute the inclination of the axis of the top as a function of the geographical latitude  $\phi$  and the mass  $m$  of the counterweight. Remember that for any given location the altitude of the pole is equal to the latitude.

*Ex. 53.* A projectile fired from a rifled gun is given a spin about its axis; the projection of the trajectory upon a horizontal plane deviates from a straight line. How may this be explained qualitatively, and what is the direction of deviation for a shell having right-handed spin, i.e. clockwise rotation when looking along the tangent to the orbit? Air resistance tends to raise the nose of the projectile.



## CHAPTER VIII

### ELASTICITY: THE MECHANICS OF DEFORMABLE SOLIDS

#### 1. The Geometry of Small Displacements.

The ideal rigid body discussed in the previous chapter represents an abstraction, as is well known. Actually, the parts of a physical solid body, e.g. a bar of iron, are not absolutely rigidly interconnected, but suffer small displacements relative to each other under the influence of forces. These deformations are, in general, so small that they can be neglected in a large number of cases. On the other hand, the phenomena governed by elastic properties, e.g. the vibrations of a spring, are often of great importance, so that elastic processes offer a wide field for discussion in mathematical physics, a field which has been extensively cultivated, for practical reasons. We can take up here only a few of the chief problems and their solutions.

In the light of our present concepts regarding the structure of matter, we can understand, in many cases, that considerable forces are necessary to deform a solid body. In the absence of external forces, the solid is in a state of internal equilibrium, in which its fundamental constituents (atoms or molecules) are at such distances apart that the forces between them are in equilibrium. Every change of these distances, caused by external forces, immediately causes the attractive or repulsive forces to preponderate; their net result, which is either an internal pressure or tension, balances the external forces. We may construct a crude model by imagining the elementary particles to be connected by spiral springs. This model shows us how a local deformation must be transmitted to all parts of the body. The mathematical computation of the deformation resulting from given forces is a problem in *Elastostatics*, while *Elastodynamics* deals with the changes of elastic disturbances in time. For example, if, in the above model, we displace several "atoms" from their rest positions and then release them, they will move beyond those points in returning; and this vibration will spread throughout the whole body, on account of the connexions between the atoms. However, for purposes of calculation we shall not make use of the atomic picture, but shall treat the phenomena from a macroscopic standpoint, assuming that the space is continuously filled with matter. This means that the portions of

space considered, and also their volume increments, are so large that they contain a great number of atoms.\*

We now wish to investigate the geometric-kinematic relationships for infinitely small deformations or *strains*. For this purpose we consider a region around a point  $P_0$ , whose dimensions are so small that the squares of the distances of all its points  $P$  from  $P_0$  may be neglected. On account of the strain, let each point experience a displacement  $\mathbf{s} = u\mathbf{i} + v\mathbf{j} + w\mathbf{k}$ , which we likewise consider to be small. Moreover, let the displacement vary so little from place to place that we are justified in dropping products of the derivatives  $\partial u/\partial x$ ,  $\partial u/\partial y$ ,  $\partial u/\partial z$ ,  $\partial v/\partial x$ , . . . If the point  $P_0$  experiences a displacement  $\mathbf{s}_0$ , then another point  $P$ , whose position relative to  $P_0$  is given by the radius vector  $\mathbf{r}$  drawn from  $P_0$ , undergoes a displacement

$$\mathbf{s} = \mathbf{s}_0 + \mathbf{r} \text{ grad} \cdot \mathbf{s}. \quad . \quad . \quad . \quad . \quad . \quad (1)$$

If we denote the components of  $\mathbf{s}$  by  $u, v, w$ , the above equation may be written

$$\left. \begin{aligned} u &= u_0 + \mathbf{r} \text{ grad } u = u_0 + \left(\frac{\partial u}{\partial x}\right)_0 x + \left(\frac{\partial u}{\partial y}\right)_0 y + \left(\frac{\partial u}{\partial z}\right)_0 z \\ v &= v_0 + \mathbf{r} \text{ grad } v = v_0 + \left(\frac{\partial v}{\partial x}\right)_0 x + \left(\frac{\partial v}{\partial y}\right)_0 y + \left(\frac{\partial v}{\partial z}\right)_0 z \\ w &= w_0 + \mathbf{r} \text{ grad } w = w_0 + \left(\frac{\partial w}{\partial x}\right)_0 x + \left(\frac{\partial w}{\partial y}\right)_0 y + \left(\frac{\partial w}{\partial z}\right)_0 z \end{aligned} \right\} \quad . \quad (2)$$

It is intended that the symbol

$$\left(\frac{\partial u}{\partial x}\right)_0$$

should express the fact that, on account of the general hypothesis concerning the order of magnitude of the displacement components and of the volume under consideration, we may use the values of the derivatives at  $P_0$  in carrying out the computation. Naturally, we rule out discontinuities, such as fractures or cracks in the body.

Our equations give a linear relationship between the components of  $\mathbf{s}$  and those of  $\mathbf{r}$ . More important, however, is a related quantity: the connexion between an initial vector  $\mathbf{r} = \vec{P_0 P}$  and the corresponding vector  $\mathbf{r}' = \vec{P'_0 P'}$  after displacement. This connexion may be looked upon as a mathematical transformation of the space

\* See pp. 2, 3.

about  $P_0$ , determined by the deformation. From fig. 1 it follows that

$$\mathbf{r}' = \mathbf{r} + (\mathbf{s} - \mathbf{s}_0) = \mathbf{r} + \mathbf{r} \text{ grad} \cdot \mathbf{s} \text{ or } \mathbf{r}' - \mathbf{r} = \mathbf{r} \text{ grad} \cdot \mathbf{s}, \quad (3)$$

or, in components,\*

$$\left. \begin{aligned} x' &= x \left( 1 + \frac{\partial u}{\partial x} \right) + y \frac{\partial u}{\partial y} + z \frac{\partial u}{\partial z} \\ y' &= x \frac{\partial v}{\partial x} + y \left( 1 + \frac{\partial v}{\partial y} \right) + z \frac{\partial v}{\partial z} \\ z' &= x \frac{\partial w}{\partial x} + y \frac{\partial w}{\partial y} + z \left( 1 + \frac{\partial w}{\partial z} \right) \end{aligned} \right\} \dots \dots \dots (4)$$

Thus this transformation, also, is linear. Since the conversion of  $\mathbf{r}$  to  $\mathbf{r}'$  is independent of the co-ordinate system we select, the coefficients of (4) represent the nine components of a tensor.

The correspondence between the space about  $P_0$  and that about  $P'_0$  is that of an affine transformation of space, where—by definition—straight lines are transformed into straight lines, and thus planes pass over into planes, and parallelism is retained. This is immediately evident

in the vector form: A set of parallel lines in the direction of  $\mathbf{t}$  may be represented by means of a variable parameter  $\lambda$  in the form

$$\mathbf{r} = \mathbf{r}_i + \lambda \mathbf{t}, \quad \dots \dots \dots (5)$$

so that one straight line of the family corresponds to each  $\mathbf{r}_i$ . If this relationship is substituted in (3), we obtain

$$\mathbf{r}' = \mathbf{r}_i + \lambda \mathbf{t} + (\mathbf{r}_i + \lambda \mathbf{t}) \text{ grad} \cdot \mathbf{s} = \mathbf{r}_i + \mathbf{r}_i \text{ grad} \cdot \mathbf{s} + \lambda (\mathbf{t} + \mathbf{t} \text{ grad} \cdot \mathbf{s}). \quad (6)$$

This, however, is again the equation of a family of parallel lines with

$$\mathbf{r}'_i = \mathbf{r}_i + \mathbf{r}_i \text{ grad} \cdot \mathbf{s}, \quad \mathbf{t}' = \mathbf{t} + \mathbf{t} \text{ grad} \cdot \mathbf{s}. \quad \dots \dots (7)$$

The transformation (4) may include still another part, which represents the rotation of the neighbouring space as a rigid entity. In this case, the vector  $\mathbf{r}$  is altered, as before, but no internal elastic forces arise, since there is no actual deformation of the volume under consideration. If we wish to find the forces attributable to changes in the distances between particles, we must deduct from (4) the part

\* For simplicity, the subscripts of the derivatives are now omitted.

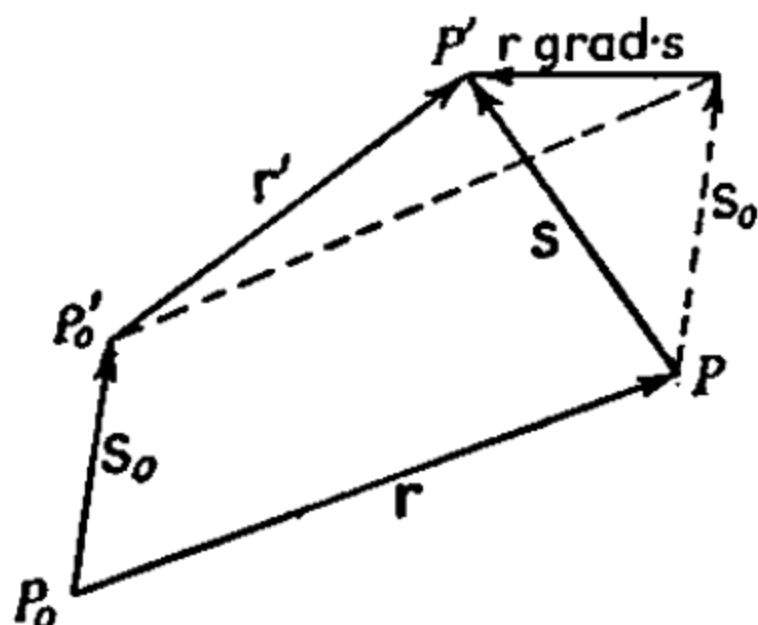


Fig. 1



concerned with pure rotation. This may be done in the following manner:

On p. 140 it was shown that the velocity of any point of a rigid body rotating with angular velocity  $\omega$  is given by

$$\mathbf{v} = [\omega \mathbf{r}]. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

Forming the curl, we get

$$\begin{aligned} \text{curl } \mathbf{v} &= [\nabla[\omega \mathbf{r}]] = \omega \cdot \nabla \mathbf{r} - \omega \nabla \cdot \mathbf{r} \\ &= \omega \text{ div } \mathbf{r} - \omega \text{ grad } \cdot \mathbf{r} = 3\omega - \omega = 2\omega. \quad . \quad (9) \end{aligned}$$

Thus, for a given velocity field, curl  $\mathbf{v}$  represents a vector equal to *twice* the angular velocity vector.\*

If a rigid body is thought of as rotating for a short time  $dt$ , it is found that the change in the radius vector  $\dagger$  caused by this rotation is

$$d_{\text{rot}} \mathbf{r} = [\omega \mathbf{r}] dt = \frac{1}{2} \left[ \text{rot } \frac{d\mathbf{s}}{dt} dt, \mathbf{r} \right] = \frac{1}{2} [\text{rot } \mathbf{s}, \mathbf{r}]. \quad . \quad (10)$$

The difference between the vector  $\mathbf{r} + d_{\text{rot}} \mathbf{r}$  and the vector  $\mathbf{r}'$  as specified by (3) represents the change in  $\mathbf{r}$  due to pure deformation, which we denote by  $d_{\text{def}} \mathbf{r}$ :

$$d_{\text{def}} \mathbf{r} = \mathbf{r}' - \mathbf{r} - d_{\text{rot}} \mathbf{r}. \quad . \quad . \quad . \quad . \quad . \quad (11)$$

From the deformation alone we then have the relation

$$\mathbf{r}'' = \mathbf{r} + d_{\text{def}} \mathbf{r} = \mathbf{r} + \mathbf{r} \text{ grad } \cdot \mathbf{s} - \frac{1}{2} [\text{curl } \mathbf{s}, \mathbf{r}], \quad . \quad (12)$$

or, in Cartesian form,

$$\left. \begin{aligned} x'' &= \left(1 + \frac{\partial u}{\partial x}\right)x + \frac{1}{2} \left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}\right)y + \frac{1}{2} \left(\frac{\partial w}{\partial x} + \frac{\partial u}{\partial z}\right)z, \\ y'' &= \frac{1}{2} \left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}\right)x + \left(1 + \frac{\partial v}{\partial y}\right)y + \frac{1}{2} \left(\frac{\partial v}{\partial z} + \frac{\partial w}{\partial y}\right)z, \\ z'' &= \frac{1}{2} \left(\frac{\partial w}{\partial x} + \frac{\partial u}{\partial z}\right)x + \frac{1}{2} \left(\frac{\partial v}{\partial z} + \frac{\partial w}{\partial y}\right)y + \left(1 + \frac{\partial w}{\partial z}\right)z. \end{aligned} \right\} \quad (12a)$$

The coefficients, in (12a), now form a *symmetric* tensor, which may be

\* This is still true if the whole body is subject, in addition, to a translation  $\mathbf{v}_T$ , for since  $\mathbf{v}_T$  is the same for all particles, it drops out in forming the curl.

$\dagger$  Actually, one should write curl  $d\mathbf{s}$ , but according to the original assumptions all displacements are assumed to be small quantities.

visualized by means of an ellipsoid, as at p. 37. This ellipsoid, which has the equation

$$\left(1 + \frac{\partial u}{\partial x}\right)x^2 + \left(1 + \frac{\partial v}{\partial y}\right)y^2 + \left(1 + \frac{\partial w}{\partial z}\right)z^2 + \left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}\right)xy + \left(\frac{\partial v}{\partial z} + \frac{\partial w}{\partial y}\right)yz + \left(\frac{\partial w}{\partial x} + \frac{\partial u}{\partial z}\right)zx = 1, \quad \dots \dots \dots (13)$$

may be called the ellipsoid of dilatation, or ellipsoid of strain. Its principal axes are called the principal axes of strain. As shown at p. 37, the direction of  $\mathbf{r}''$  is that of the normal to the ellipsoid drawn through the end point of  $\mathbf{r}$ . The two vectors have the same direction for the ends of the principal axes. This means that only a segment having this direction maintains its original direction. *Only those segments lying in the directions of the axes of the strain ellipsoid experience pure extension without change of direction.*

The coefficients

$$1 + \frac{\partial u}{\partial x}, \quad \frac{1}{2}\left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}\right), \quad \frac{1}{2}\left(\frac{\partial w}{\partial x} + \frac{\partial u}{\partial z}\right), \text{ \&c.,}$$

have a definite physical meaning: if we consider the termini of the unit vectors along the axes, for the undeformed body, then, corresponding to the vector  $\mathbf{i}$ , we have, by equations (12a),

$$\mathbf{r}_i'' = \left(1 + \frac{\partial u}{\partial x}\right)\mathbf{i} + \frac{1}{2}\left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}\right)\mathbf{j} + \frac{1}{2}\left(\frac{\partial w}{\partial x} + \frac{\partial u}{\partial z}\right)\mathbf{k}, \quad (14)$$

and for the vector  $\mathbf{j}$ ,

$$\mathbf{r}_j'' = \frac{1}{2}\left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}\right)\mathbf{i} + \left(1 + \frac{\partial v}{\partial y}\right)\mathbf{j} + \frac{1}{2}\left(\frac{\partial v}{\partial z} + \frac{\partial w}{\partial y}\right)\mathbf{k}. \quad (15)$$

Neglecting all terms which may be considered small according to the assumptions made above, we have

$$\frac{|\mathbf{r}_i''| - |\mathbf{i}|}{|\mathbf{i}|} = \frac{\sqrt{\mathbf{r}_i''^2} - 1}{1} = \frac{\partial u}{\partial x}, \quad \dots \dots \dots (16)$$

$$\frac{|\mathbf{r}_j''| - |\mathbf{j}|}{|\mathbf{j}|} = \frac{\partial v}{\partial y}. \quad \dots \dots \dots (17)$$

This means that the quantities

$$\frac{\partial u}{\partial x}, \quad \frac{\partial v}{\partial y}, \quad \frac{\partial w}{\partial z}$$

give the relative changes in length of segments lying in the  $x$ -,  $y$ -, and  $z$ -directions respectively.

In the undeformed state of the body, the vectors  $i$  and  $j$  are perpendicular to each other; after deformation, they form an angle which we may designate by  $\pi/2 - \delta_{ij}$ , since it differs but little from  $\pi/2$ . Then,

$$\cos\left(\frac{\pi}{2} - \delta_{ij}\right) = \sin \delta_{ij} = \frac{r_i'' r_j''}{|r_i''| |r_j''|}. \quad . \quad . \quad (18)$$

If we use the values of  $r_i''$  and  $r_j''$  from equations (14) and (15), and if we neglect all products of the derivatives, we obtain

$$\sin \delta_{ij} \approx \delta_{ij} = \left(\frac{\partial u}{\partial y} + \frac{\partial v}{\partial x}\right). \quad . \quad . \quad . \quad (19)$$

Thus the tensor component  $2T_{12}$  signifies the change in angle, between two segments having the directions of the  $x$ - and  $y$ -axes, resulting from the deformation. Similar results hold for the other components. If the axes have the directions of the principal axes of the tensor ellipsoid, then the terms  $T_{12}$ ,  $T_{23}$ ,  $T_{31}$  are absent from the co-ordinate representation. This means that the axes of the strain ellipsoid remain perpendicular to each other, which we already inferred above from the fact that segments in these directions suffer pure contraction or expansion.

To calculate the change in volume per unit volume (the so-called "cubical dilatation"), we need only find the difference in the volume of the parallelepiped formed by the vectors  $i$ ,  $j$ ,  $k$ , before and after deformation, i.e. the difference  $[r_i'' r_j'' r_k''] - 1$ . On writing out the determinant we see immediately that this expression, to a first approximation as before, gives

$$\frac{V'' - V}{V} = \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} = \text{div } s. \quad . \quad . \quad (20)$$

It is more convenient, for what follows, to consider the strains themselves, i.e. the components of the vector  $r'' - r$ . These are obtained at once from equations (12a) (p. 164) by bringing the components of  $r$  over to the left side. Introduce the following notation for the derivatives and their combinations, the physical meaning of which was given above:

$$\left. \begin{aligned} e_{11} &= \frac{\partial u}{\partial x} & e_{12} &= \frac{\partial u}{\partial y} + \frac{\partial v}{\partial x} = e_{21} \\ e_{22} &= \frac{\partial v}{\partial y} & e_{23} &= \frac{\partial v}{\partial z} + \frac{\partial w}{\partial y} = e_{32} \\ e_{33} &= \frac{\partial w}{\partial z} & e_{31} &= \frac{\partial w}{\partial x} + \frac{\partial u}{\partial z} = e_{13} \end{aligned} \right\} . \quad . \quad (21)$$



We then may write, by (12a),

$$\left. \begin{aligned} x'' - x &= e_{11}x + \frac{1}{2}e_{12}y + \frac{1}{2}e_{13}z \\ y'' - y &= \frac{1}{2}e_{21}x + e_{22}y + \frac{1}{2}e_{23}z \\ z'' - z &= \frac{1}{2}e_{31}x + \frac{1}{2}e_{32}y + e_{33}z \end{aligned} \right\} \quad \cdot \quad \cdot \quad \cdot \quad (22)$$

The tensor ellipsoid corresponding to this symmetric vector function has the same axial directions as the ellipsoid of strain (13) (p. 165). This may be shown readily by calculation, but also can be seen at once in this way: if the vector  $\mathbf{r}''$  is to have the direction of  $\mathbf{r}$  for the axis, then this must also be true for the difference vector  $\mathbf{r}'' - \mathbf{r}$ . The mixed tensor components vanish when referred to the principal axes. Denoting by accents the co-ordinates referred to this system, we have remaining,

$$e_1 = \frac{\partial u'}{\partial x'}, \quad e_2 = \frac{\partial v'}{\partial y'}, \quad e_3 = \frac{\partial w'}{\partial z'}. \quad \cdot \quad \cdot \quad (23)$$

The quantities  $e_1, e_2, e_3$  are called the *principal extensions*.

*Ex. 54.* Compute the relative change in length of a segment whose direction cosines, referred to the axes of the ellipsoid of dilatation, are  $\alpha_1, \alpha_2, \alpha_3$ .

## 2. State of Stress of a Body under Strain.

If we consider an element of volume  $d\tau$ , of mass  $dm$ , in the interior of a strained elastic body, we recognize that two kinds of force are operating. First, there are forces like that of gravity, applied to all the individual constituents, whether they be in the interior of the body or on the bounding surfaces. If the body is being accelerated, the "inertial resistance"

$$-dm \left( \frac{d^2 \mathbf{r}}{dt^2} \right)$$

is also a force of this type. But, in addition to forces of this kind, there are forces operating on the surface particles which are due to the presence of the contiguous particles. These correspond to the springs in the model described on p. 161. Since these forces operate between adjacent particles on either side of the bounding surface, they are not proportional to the volume of the element, but to the area of its surface. The value of the force acting upon an element of surface, referred to unit area, is called the *unital stress*, or simply the *stress*,  $\mathbf{P}$ . The direction of this force need not by any means be normal to the surface—in general it is composed of a normal stress (pressure or tension) and a tangential (or shearing) stress. We can show, however, that the stress  $\mathbf{P}$  on any surface element may be found in terms of the normal direction, provided that the stresses are known at this

point for three mutually perpendicular directions, which we choose as axes of a rectangular co-ordinate system. Thus, we take the volume element to be a small tetrahedron, of which three faces— $dS_x$ ,  $dS_y$ ,  $dS_z$ —lie in the co-ordinate planes, while the fourth— $dS$ —has the outward directed normal  $\mathbf{n}$  (fig. 2). Let the stress on  $dS$  be  $\mathbf{P}$ . In addition, call the stress acting on the surface whose normal is along the  $x$ -axis  $\mathbf{P}_1$ , and the stresses on the faces perpendicular to the  $y$ - and  $z$ -axes  $\mathbf{P}_2$  and  $\mathbf{P}_3$  respectively. The vectors belonging to the faces which lie in the co-ordinate planes have the directions of the *negative*  $x$ -,  $y$ - and  $z$ -axes. The stresses  $\mathbf{P}_1$ ,  $\mathbf{P}_2$ ,  $\mathbf{P}_3$  therefore denote the stresses exerted by the faces of the tetrahedron on the adjacent surface elements, whose normals are along the *positive* axes. According to Newton's Third Law, the stresses exerted by the neighbouring surfaces on the tetrahedron are then  $-\mathbf{P}_1$ ,  $-\mathbf{P}_2$ ,  $-\mathbf{P}_3$ . If we now formulate the conditions of equilibrium for our infinitely small tetrahedron, we need consider only the forces proportional to the areas, since the forces proportional to the volume vanish to a higher order as we diminish the dimensions of the element. Then the first condition of equilibrium—the vanishing of the resultant force—is

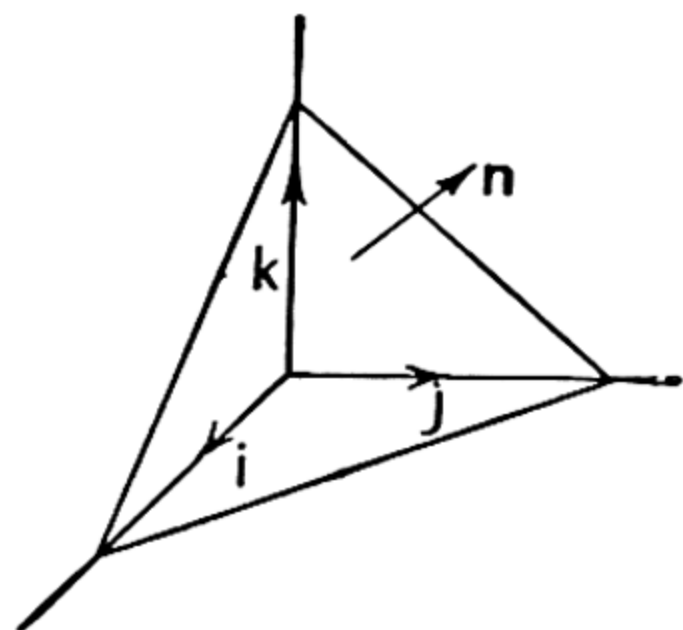


Fig. 2

$$\mathbf{P}dS - \mathbf{P}_1dS_x - \mathbf{P}_2dS_y - \mathbf{P}_3dS_z = 0. \quad . \quad . \quad (24)$$

Now 
$$\mathbf{n}dS = i dS_x + j dS_y + k dS_z, \quad . \quad . \quad . \quad (25)$$

that is,

$$dS_x = dS(\mathbf{n}i) = dS \cos \alpha; \quad dS_y = dS \cos \beta; \quad dS_z = dS \cos \gamma, \quad (26)$$

whence 
$$\mathbf{P} = \mathbf{P}_1 \cos \alpha + \mathbf{P}_2 \cos \beta + \mathbf{P}_3 \cos \gamma. \quad . \quad . \quad . \quad (27)$$

Denoting the  $x$ -component of  $\mathbf{P}_1$  by  $P_{11}$ , that of  $\mathbf{P}_2$  by  $P_{21}$ , the  $y$ -component of  $\mathbf{P}_1$  by  $P_{12}$ , &c., we obtain the component representation

$$\left. \begin{aligned} P_x &= P_{11} \cos \alpha + P_{21} \cos \beta + P_{31} \cos \gamma \\ P_y &= P_{12} \cos \alpha + P_{22} \cos \beta + P_{32} \cos \gamma \\ P_z &= P_{13} \cos \alpha + P_{23} \cos \beta + P_{33} \cos \gamma \end{aligned} \right\} \quad . \quad . \quad (28)$$

This equation gives every normal direction  $\mathbf{n}$  a corresponding stress vector  $\mathbf{P}$ .

The meaning of the coefficients  $P_{ik}$  may be seen by considering a cube of edge unity, whose faces are parallel to the co-ordinate planes. Then, for example, for one of the faces parallel to the  $yz$ -plane,

$$\cos \alpha = 1, \quad \cos \beta = \cos \gamma = 0,$$

i.e. 
$$P_x = P_{11}, \quad P_y = P_{12}, \quad P_z = P_{13},$$

so that the normal component  $P_x$  is represented by  $P_{11}$ , and the two components lying in the surface are represented by  $P_{12}$  and  $P_{13}$ . Here, then, the coefficients having two equal subscripts give the normal pressures or tensions, while those with mixed subscripts represent the tangential stresses.

The coefficients again represent the components of a tensor, since the connexion between  $\mathbf{P}$  and  $\mathbf{n}$  must, of course, be independent of the particular co-ordinate system in which the vectors are resolved, and must always lead from a given vector  $\mathbf{n}$  to the same vector  $\mathbf{P}$ . The tensor defined by the quantities  $P_{ik}$  is symmetric, as will be shown in the following section. It may therefore be represented again by an ellipsoid, called the *ellipsoid of stress*. If the co-ordinate axes are taken along the principal axes of the ellipsoid, the mixed tensor components drop out, and we have only the three coefficients  $P_1, P_2, P_3$ , corresponding to the diagonal terms of the matrix of coefficients. These three quantities are known as the *principal stresses*. The relationship between the components of the stress acting on a surface element normal to  $\mathbf{n}$  and the principal stresses is very simple when expressed in this system of co-ordinates. Denoting the quantities referred to this system by primes, we have

$$\left. \begin{aligned} P_{x'} &= P_1 \cos \alpha' \\ P_{y'} &= P_2 \cos \beta' \\ P_{z'} &= P_3 \cos \gamma' \end{aligned} \right\} \dots \dots \dots (29)$$

### 3. The Conditions of Equilibrium of an Elastic Body.

Consider the equilibrium of the forces applied to a finite volume of an elastic body. If we denote the resultant of the forces proportional to the volume—the body forces—acting on unit volume by  $\mathbf{g}$ , and apply the fundamental equations of the mechanics of systems of particles, we obtain

$$\int \mathbf{g} d\tau + \oint \mathbf{P} dS = 0. \dots \dots \dots (30)$$

$$\int [\mathbf{r} \mathbf{g}] d\tau + \oint [\mathbf{r} \mathbf{P}] dS = 0. \dots \dots \dots (31)$$

The surface integral is to be taken over the bounding surface of the volume under consideration. In adding the stresses acting upon the several volume elements, the internal stresses on the boundary between two elements cancel out, since these stresses are applied in opposite directions on the two sides of such a surface of separation.

If we take the  $x$ -component of equation (30) we have on the left

$$\int g_x d\tau + \oint P_x dS = \int g_x d\tau + \oint (P_{11} \cos \alpha + P_{21} \cos \beta + P_{31} \cos \gamma) dS. (32)$$

The surface integral may be brought into the form appearing in Gauss's



Theorem and then transformed into a volume integral if we introduce another vector  $\bar{\mathbf{P}}_1$ :

$$\bar{\mathbf{P}}_1 = P_{11}\mathbf{i} + P_{21}\mathbf{j} + P_{31}\mathbf{k}. \quad . \quad . \quad . \quad (33)$$

We shall show, below, that the tensor  $P_{ik}$  is symmetric, so that  $\bar{\mathbf{P}}_1$  is identical with  $\mathbf{P}_1$ . Using this new vector,

$$\oint P_x dS = \oint (\bar{\mathbf{P}}_1 \mathbf{n}) dS = \oint (\bar{\mathbf{P}}_1 d\mathbf{S}) = \int \text{div } \bar{\mathbf{P}}_1 d\tau. \quad (34)$$

Since the condition of equilibrium must hold for every volume, we have, writing the analogous equations for the  $y$ - and  $z$ -components:

$$\left. \begin{aligned} g_x + \text{div } \bar{\mathbf{P}}_1 &= g_x + \frac{\partial P_{11}}{\partial x} + \frac{\partial P_{21}}{\partial y} + \frac{\partial P_{31}}{\partial z} = 0 \\ g_y + \text{div } \bar{\mathbf{P}}_2 &= g_y + \frac{\partial P_{12}}{\partial x} + \frac{\partial P_{22}}{\partial y} + \frac{\partial P_{32}}{\partial z} = 0 \\ g_z + \text{div } \bar{\mathbf{P}}_3 &= g_z + \frac{\partial P_{13}}{\partial x} + \frac{\partial P_{23}}{\partial y} + \frac{\partial P_{33}}{\partial z} = 0 \end{aligned} \right\}. \quad (35)$$

Again, the  $x$ -component of the equation of moments becomes

$$\int (yg_z - zg_y) d\tau + \oint (yP_z - zP_y) dS = 0. \quad . \quad . \quad (36)$$

Just as before, we may write

$$\oint (yP_z - zP_y) dS = \oint (y\bar{\mathbf{P}}_3 - z\bar{\mathbf{P}}_2) d\mathbf{S} = \int (\text{div } y\bar{\mathbf{P}}_3 - \text{div } z\bar{\mathbf{P}}_2) d\tau. \quad (37)$$

Now

$$\text{div } y\bar{\mathbf{P}}_3 = y \text{div } \bar{\mathbf{P}}_3 + \bar{\mathbf{P}}_3 (\text{grad } y) = y \text{div } \bar{\mathbf{P}}_3 + \bar{\mathbf{P}}_3 \mathbf{j} = y \text{div } \bar{\mathbf{P}}_3 + P_{23}, \quad (38)$$

so that the  $x$ -component of the equation of moments becomes

$$y(g_z + \text{div } \bar{\mathbf{P}}_3) - z(g_y + \text{div } \bar{\mathbf{P}}_2) - P_{32} + P_{23} = 0. \quad (39)$$

But, according to the first condition of equilibrium, the bracketed expressions are zero, and so we arrive at the proof of the symmetry of the stress tensor:

$$P_{32} = P_{23}. \quad . \quad . \quad . \quad . \quad . \quad (40)$$

Similarly, from the other two component equations of (31),

$$P_{13} = P_{31}; \quad P_{21} = P_{12}. \quad . \quad . \quad . \quad . \quad (40')$$

In addition to the conditions for equilibrium of internal forces we have those for the surface forces. If the stresses  $\mathbf{P}^e$  on the surface elements, caused by the external forces, are known, then equations (28) (p. 168) hold again, and we have

$$\left. \begin{aligned} P_x^e &= P_{11} \cos \alpha + P_{21} \cos \beta + P_{31} \cos \gamma \\ P_y^e &= P_{12} \cos \alpha + P_{22} \cos \beta + P_{32} \cos \gamma \\ P_z^e &= P_{13} \cos \alpha + P_{23} \cos \beta + P_{33} \cos \gamma \end{aligned} \right\}. \quad . \quad . \quad (41)$$

## 4. Relations between the Strain Tensor and the Stress Tensor.

The conditions of equilibrium (35) are not yet sufficient for the answer to the question "What equilibrium state of strain results when given external forces are applied?" We still lack the principal thing—the connexion between the stress and the strain. Since each consists of six components, we must expect, in the most general case, that each component of the strain tensor is a function of all the components of the stress tensor. Assuming a linear relationship, this leads to a system of equations with thirty-six coefficients. This number reduces very considerably, according to the symmetry properties of the substance under consideration. If we consider only isotropic substances (i.e. not crystals), then, as we shall see, two constants are sufficient. For an isotropic substance, the axes of the stress ellipsoid must coincide with those of the strain ellipsoid, for the former give the directions for which the stresses are normal to the surface; and for an isotropic body, only contractions or extensions, but not lateral displacements, can occur in these directions. Now, according to § 1 (p. 165), this is the essential property of the axes of the strain ellipsoid. We shall therefore express the connexion between the stress and strain tensors in this uniquely indicated system of co-ordinates. The foundation of the so-called classical theory of elasticity is *Hooke's Law*, which states that the normal stress is proportional to the elongation per unit length ("ut tensio, sic vis"). It is understood that we limit ourselves to small strains, for—according to experience—Hooke's Law is no longer valid for larger deformations. If we write this law for the axis 1 in the form  $P_1 = Ee_1$ , then  $E$  is the linear modulus of elasticity, or *Young's modulus*, for the material in question. We know from experience that a body undergoes a contraction of its transverse sections when stretched in the direction of its length. The ratio of the transverse contraction—i.e. the change in length per unit length of a transverse segment—to the change in length per unit length of the body is, according to experiment, constant within the limits where Hooke's Law holds. This ratio is called *Poisson's Ratio*  $\sigma$ . That is, if stresses  $P_1, P_2, P_3$  act in the directions of the axes, then the tension  $P_2$  causes a *contraction* along the axis 1. This relative contraction is given by  $\sigma e_2$ , and since  $P_2 = Ee_2$ , it amounts to  $(\sigma/E)P_2$ . Similarly, the tension  $P_3$  causes a contraction of amount  $(\sigma/E)P_3$ . The total relative change in length is thus given by

$$\left. \begin{aligned} e_1 &= \frac{1}{E} \{P_1 - \sigma(P_2 + P_3)\} = \frac{1+\sigma}{E} \left\{ P_1 - \frac{\sigma}{1+\sigma} (P_1 + P_2 + P_3) \right\} \\ e_2 &= \frac{1}{E} \{P_2 - \sigma(P_3 + P_1)\} = \frac{1+\sigma}{E} \left\{ P_2 - \frac{\sigma}{1+\sigma} (P_1 + P_2 + P_3) \right\} \\ e_3 &= \frac{1}{E} \{P_3 - \sigma(P_1 + P_2)\} = \frac{1+\sigma}{E} \left\{ P_3 - \frac{\sigma}{1+\sigma} (P_1 + P_2 + P_3) \right\} \end{aligned} \right\}. \quad (42)$$

It remains only to transform these equations, derived in this way for the special co-ordinate system of the principal axes of the ellipsoids, to an arbitrary co-ordinate system, since, in general, the principal axes for a strained body change in direction from place to place, and thus are not suited to describe what occurs. Let the axis 1 form an angle  $(1, x)$  with the  $x$ -axis, an angle  $(1, y)$  with the  $y$ -axis, and an angle  $(1, z)$  with the  $z$ -axis, and let the angles between the axes 2 and 3 and the new axes be denoted in similar manner. Further, let us characterize the principal axes 1, 2, 3 by the unit vectors  $\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3$ , the new axes by the unit vectors  $\mathbf{i}, \mathbf{j}, \mathbf{k}$ . Then

$$\left. \begin{aligned} \mathbf{i} &= \mathbf{c}_1 \cos(1, x) + \mathbf{c}_2 \cos(2, x) + \mathbf{c}_3 \cos(3, x) \\ \mathbf{j} &= \mathbf{c}_1 \cos(1, y) + \mathbf{c}_2 \cos(2, y) + \mathbf{c}_3 \cos(3, y) \\ \mathbf{k} &= \mathbf{c}_1 \cos(1, z) + \mathbf{c}_2 \cos(2, z) + \mathbf{c}_3 \cos(3, z) \end{aligned} \right\}. \quad (43)$$

By the rule for transformation of tensor components proved on p. 34 we obtain

$$\left. \begin{aligned} e_{11} &= e_1 \cos^2(1, x) + e_2 \cos^2(2, x) + e_3 \cos^2(3, x) \\ e_{22} &= e_1 \cos^2(1, y) + e_2 \cos^2(2, y) + e_3 \cos^2(3, y) \\ e_{33} &= e_1 \cos^2(1, z) + e_2 \cos^2(2, z) + e_3 \cos^2(3, z) \\ \frac{1}{2}e_{12} = \frac{1}{2}e_{21} &= e_1 \cos(1, x) \cos(1, y) + e_2 \cos(2, x) \cos(2, y) \\ &\quad + e_3 \cos(3, x) \cos(3, y) \\ \frac{1}{2}e_{23} = \frac{1}{2}e_{32} &= e_1 \cos(1, y) \cos(1, z) + e_2 \cos(2, y) \cos(2, z) \\ &\quad + e_3 \cos(3, y) \cos(3, z) \\ \frac{1}{2}e_{31} = \frac{1}{2}e_{13} &= e_1 \cos(1, z) \cos(1, x) + e_2 \cos(2, z) \cos(2, x) \\ &\quad + e_3 \cos(3, z) \cos(3, x) \end{aligned} \right\}. \quad (44)$$

It is evident from these expressions that the sum

$$e_{11} + e_{22} + e_{33} = \text{div } \mathbf{s}$$

is equal to  $e_1 + e_2 + e_3$ , independent of the co-ordinate system. Such quantities are called invariants of a tensor. A second invariant which will be useful later (cf. p. 176) is the expression

$$e_{11}e_{22} + e_{22}e_{33} + e_{33}e_{11} - \frac{1}{4}(e_{12}^2 + e_{23}^2 + e_{31}^2) = e_1e_2 + e_2e_3 + e_3e_1. \quad (45)$$

The same transformation formulæ hold for the components of the stress tensor  $P_{ik}$ . Thus, expressing  $e_1, e_2, e_3$  in (44) in terms of  $P_1, P_2, P_3$ , using equations (42), and also the relationships

$$P_1 + P_2 + P_3 = P_{11} + P_{22} + P_{33}$$

and  $\cos^2(1, x) + \cos^2(2, x) + \cos^2(3, x) = 1,$

$$\begin{aligned} \cos(1, x) \cos(1, y) + \cos(2, x) \cos(2, y) \\ + \cos(3, x) \cos(3, y) = 0, \text{ \&c.,} \end{aligned}$$





in volume,  $\sigma$  cannot be greater than  $1/2$ . Were  $\sigma = 1/2$ , the substance would be incompressible.

Solving the equations (46) for the components of the stress tensor, we obtain

$$\left. \begin{aligned} P_{11} &= \frac{E}{1+\sigma} \left\{ e_{11} + \frac{\sigma}{1-2\sigma} (e_{11} + e_{22} + e_{33}) \right\}, & P_{12} &= \frac{E}{2(1+\sigma)} e_{12}. \\ P_{22} &= \frac{E}{1+\sigma} \left\{ e_{22} + \frac{\sigma}{1-2\sigma} (e_{11} + e_{22} + e_{33}) \right\}, & P_{23} &= \frac{E}{2(1+\sigma)} e_{23}. \\ P_{33} &= \frac{E}{1+\sigma} \left\{ e_{33} + \frac{\sigma}{1-2\sigma} (e_{11} + e_{22} + e_{33}) \right\}, & P_{31} &= \frac{E}{2(1+\sigma)} e_{31}. \end{aligned} \right\} \quad (50)$$

If we now substitute these values of the stresses in equations (35) (p. 170), and if we again introduce the values of  $e_{ik}$  from equations (21) (p. 166), we obtain a system of differential equations containing only the displacements  $u, v, w$  and their derivatives. First, the  $x$ -component becomes

$$\begin{aligned} g_x + \frac{E}{2(1+\sigma)} \left\{ 2 \frac{\partial^2 u}{\partial x^2} + \frac{2\sigma}{1-2\sigma} \left( \frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 v}{\partial x \partial y} + \frac{\partial^2 w}{\partial x \partial z} \right) \right\} \\ + \frac{E}{2(1+\sigma)} \left\{ \frac{\partial^2 u}{\partial y^2} + \frac{\partial^2 v}{\partial x \partial y} + \frac{\partial^2 w}{\partial x \partial z} + \frac{\partial^2 u}{\partial z^2} \right\} = 0. \end{aligned}$$

This may be written

$$g_x + \frac{E}{2(1+\sigma)} \left\{ \Delta u + \frac{1}{1-2\sigma} \frac{\partial}{\partial x} \left( \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} \right) \right\} = 0. \quad (51)$$

Similarly,

$$g_y + \frac{E}{2(1+\sigma)} \left\{ \Delta v + \frac{1}{1-2\sigma} \frac{\partial}{\partial y} \left( \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} \right) \right\} = 0, \quad (51')$$

and

$$g_z + \frac{E}{2(1+\sigma)} \left\{ \Delta w + \frac{1}{1-2\sigma} \frac{\partial}{\partial z} \left( \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} \right) \right\} = 0. \quad (51'')$$

In vector form, this system becomes

$$\mathbf{g} + \frac{E}{2(1+\sigma)} \left\{ \Delta \mathbf{s} + \frac{1}{1-2\sigma} \text{grad div } \mathbf{s} \right\} = 0. \quad (52)$$

This is the general system of equations for the distortion of an isotropic, elastic substance. In addition, we have the surface boundary conditions, which may be given in the form of information concerning the displacements or the distribution of stresses on the surface.

# 5. Energy of Elastically Deformed Bodies; Elastic Potential.

Work is required to bend a bar elastically. This work remains stored in the bar in the form of potential energy as long as the deformation persists. It may be drawn upon by letting the bar go, whereupon portions of the bar attain kinetic energy and go beyond the equilibrium position. The energy of elastic deformation is potential energy of the forces operating between the particles and is distributed uniformly throughout the body. It has its greatest value at places where the deformation is greatest. In view of the continuous distribution, we look for an expression for the *energy density*—the limit of the ratio energy/volume of an elementary volume.\*

To obtain an expression for the energy density  $U$ , we start with a cube of edge equal to unity. In this case the energy *content* of the cube is equal to the energy *density*. Let the edges of the cube be in the directions of the principal axes of strain. As a result, the cube suffers only extension or compression, but not shear. In its deformed state, the face perpendicular to axis 1 has an area of  $(1 + e_2)(1 + e_3)$ , and so the force acting on it is  $P_1(1 + e_2)(1 + e_3)$ . If now the edge parallel to 1, whose deformed length is  $(1 + e_1)$ , were stretched an amount  $\delta e_1$ , the stress  $P_1$  would do an amount of work equal to

$$\delta A_1 = P_1(1 + e_2)(1 + e_3)\delta e_1.$$

(The increase in stress due to the added deformation would make a contribution which would be of only second order.) The above amount of work represents the increase in potential energy due to the assumed change in deformation. Further, the quantities  $e_2$  and  $e_3$  in the parentheses may be neglected in comparison with unity, so that the increase in potential energy due to the increase in length of the axis 1 will amount to

$$\delta U_1 = \frac{\partial U}{\partial e_1} \delta e_1 = P_1 \delta e_1, \text{ or } P_1 = \frac{\partial U}{\partial e_1}. \quad \cdot \cdot \quad (53)$$

Analogous expressions result for the other axes. Substituting the values of  $P_1, P_2, P_3$  from (42) or (50), we obtain

$$\left. \begin{aligned} P_1 &= \frac{\partial U}{\partial e_1} = \frac{E}{1 + \sigma} \left[ e_1 + \frac{\sigma}{1 - 2\sigma} (e_1 + e_2 + e_3) \right] \\ P_2 &= \frac{\partial U}{\partial e_2} = \frac{E}{1 + \sigma} \left[ e_2 + \frac{\sigma}{1 - 2\sigma} (e_1 + e_2 + e_3) \right] \\ P_3 &= \frac{\partial U}{\partial e_3} = \frac{E}{1 + \sigma} \left[ e_3 + \frac{\sigma}{1 - 2\sigma} (e_1 + e_2 + e_3) \right] \end{aligned} \right\} \quad (54)$$

\* It must be remembered, however, that according to the discussion on p. 3 the element of volume must not be made too small.



It is easy to see that these equations are satisfied by setting

$$U(e_1, e_2, e_3) = \frac{E}{2(1+\sigma)} \left[ (e_1^2 + e_2^2 + e_3^2) + \frac{\sigma}{1-2\sigma} (e_1 + e_2 + e_3)^2 \right]. \quad (55)$$

This expression, which gives the energy content as a function of the principal strains, presupposes the use of the principal axes as coordinate axes to be used at each point, and this system will change from one point to another. For this reason, it would be desirable to transform the relation to a form that holds for the whole body. To do this, we do not have to use the general transformation, since the quantities involved here may be expressed readily in terms of invariants of the tensor. The second parenthesis is such an invariant as it stands. Further, because of (45),

$$\begin{aligned} e_1^2 + e_2^2 + e_3^2 &= (e_1 + e_2 + e_3)^2 - 2(e_1 e_2 + e_2 e_3 + e_3 e_1) \\ &= (e_{11} + e_{22} + e_{33})^2 \\ &\quad - 2[e_{11}e_{22} + e_{22}e_{33} + e_{33}e_{11} - \frac{1}{4}(e_{12}^2 + e_{23}^2 + e_{31}^2)]. \end{aligned}$$

Then  $U$  will be expressed in terms of the  $e_{ik}$  as follows:

$$\begin{aligned} U(e_{ik}) &= \frac{E}{2(1+\sigma)} \left[ \frac{1-\sigma}{1-2\sigma} (e_{11} + e_{22} + e_{33})^2 - 2(e_{11}e_{22} + e_{22}e_{33} + e_{33}e_{11}) \right. \\ &\quad \left. + \frac{1}{2}(e_{12}^2 + e_{23}^2 + e_{31}^2) \right]. \quad . \quad . \quad . \quad . \quad . \quad (56) \end{aligned}$$

Comparison with (50) shows again that, in general,

$$P_{ik} = \frac{\partial U}{\partial e_{ik}}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (57)$$

Finally, the deformations in (56) may be expressed in terms of the stresses, giving a quite analogous expression which, however, will not be written down. The quantity  $U(e_{ik})$  is rightly called the *elastic potential*, since the stresses are obtainable from it by differentiation.

## 6. Elementary Treatment of the Bending of a Cantilever Beam.

The exact solution of an elastostatic problem requires the integration of equation (52), subject to the boundary conditions applying in the special case considered. In general, this procedure leads to extremely complicated calculations. For this reason, approximate solutions have been devised for many technically important problems.

Certain simplifications are made in the initial assumptions, and the results are sufficiently accurate for many purposes. As an example of such a simplified treatment, we give J. Bernoulli's theory of the bending of a beam of uniform cross-section. The Bernoulli theory is based upon the following assumptions: Imagine a certain point—say the centre of gravity—to be marked in every cross-section. The resulting straight line, which we call the axis of the beam, will be bent into a curve which lies in a vertical plane, provided that the

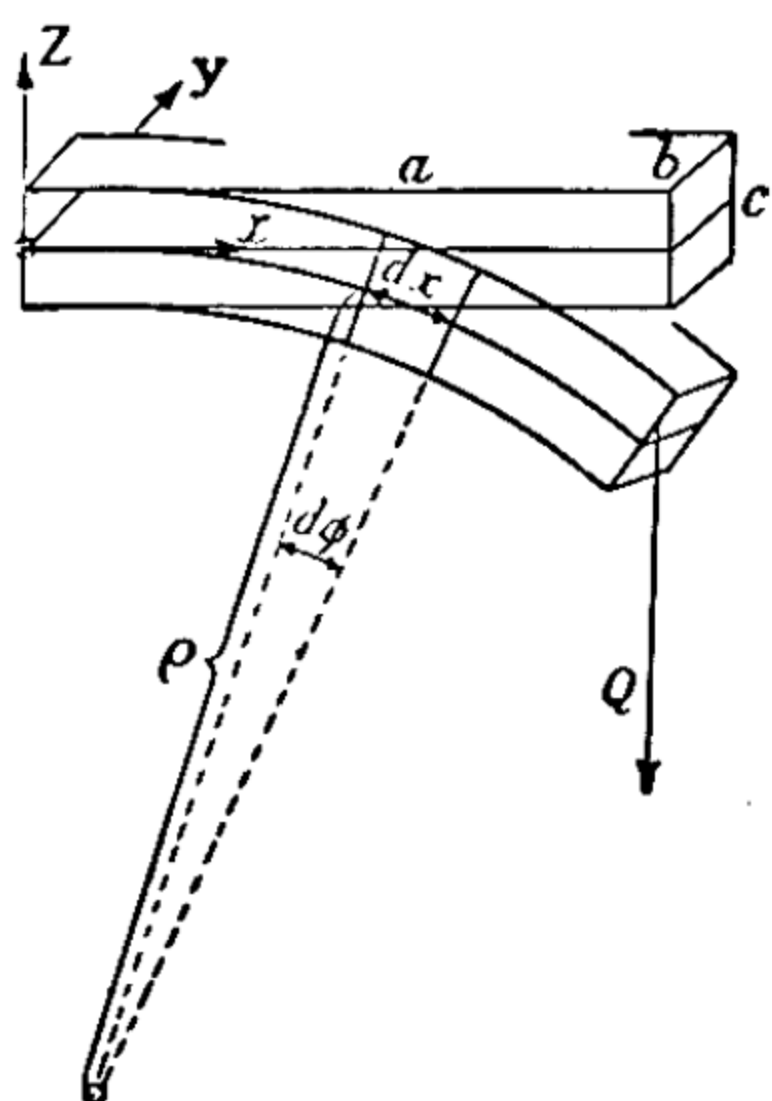


Fig. 4

bending is caused only by weights or other vertical forces. Further, imagine the beam divided into a number of thin, horizontal layers, each made up of filaments. During the bending, there will be one layer in the middle, the fibres in which are not changed in length, and which is therefore called the neutral surface. Filaments lying below the neutral surface are compressed, those above are stretched (cf. fig. 4). The flexural rigidity or stiffness of the beam is assumed to be conditioned solely by these changes in length. A further assumption of the Bernoulli theory is that plane sections normal to the axis of the beam remain plane and perpendicular to the axis after

flexure. With these assumptions, it is not difficult to calculate the elastic stresses acting upon any particular section. If we consider two neighbouring cross-sections, distant  $dx$  from one another, then these sections will form an angle  $d\phi$  with each other after bending. Measure the  $x$ -coordinate towards the right from the fixed end of the beam, and take the  $z$ -axis vertically upward from the neutral fibre, at the fixed end. The  $y$ -axis is horizontal, and directed across the beam from the nearer face. If  $\rho$  is the radius of curvature of the neutral fibre, then by fig. 4, a layer which is distant  $z$  from the neutral surface experiences a relative change in length amounting to

$$\frac{(\rho + z)d\phi - \rho d\phi}{dx} = z \frac{d\phi}{dx} = \frac{z}{\rho}, \quad \dots (58)$$

since  $d\phi/dx$  is the curvature  $1/\rho$ .

The force acting upon an element  $dzdy$  of the cross-section is, according to the definition of Young's modulus  $E$ ,

$$dF = \frac{Ez}{\rho} dz dy.$$

The total force acting upon the entire cross-section is then

$$F = \frac{E}{\rho} \iint z \, dz \, dy. \quad . \quad . \quad . \quad . \quad . \quad . \quad (59)$$

This resultant must vanish, since neither a change in cross-section nor a longitudinal displacement of any section can take place in pure bending. This means that the neutral surface always passes through the centre of gravity of the cross-section, for this is the origin for which  $\iint z \, dz \, dy$  vanishes. On the other hand, there exists a resultant *moment* of the elastic forces which tends to turn the cross-section about the line of intersection of the neutral surface with this section. In fig. 4, this moment is in the direction of the negative  $y$ -axis. Its magnitude is

$$|M| = \int z \, dF = \frac{E}{\rho} \iint z^2 \, dz \, dy = \frac{E}{\rho} I. \quad . \quad . \quad (60)$$

The geometric quantity  $\iint z^2 \, dz \, dy$  is called the moment of inertia  $I$  of the cross-section.

Let the cross-section be rectangular, as assumed in fig. 4; if we call the breadth  $b$  and the depth  $c$ , then the integration is to be extended from  $y = 0$  to  $y = b$  and from  $z = -c/2$  to  $z = +c/2$ .

The moment of the elastic forces must balance the moment of the external forces. If the bending is caused by a load  $Q$  (in force units) applied at the free end of the beam, where  $x = a$ , then the moment of this load at the point  $x$  is, since the portion  $a - x$  may be considered a rigid lever arm,

$$|M'| = Q(a - x). \quad . \quad . \quad . \quad . \quad . \quad . \quad (61)$$

The direction of this moment vector is that of the positive  $y$ -axis. The condition for equilibrium of a cross-section distant  $x$  from the fixed end is thus

$$\frac{EI}{\rho} = Q(a - x). \quad . \quad . \quad . \quad . \quad . \quad . \quad (62)$$

We can now readily obtain the equation of the neutral axis after flexure. Since the curve departs but little from the horizontal direction, we may write

$$\frac{1}{\rho} = - \frac{d^2 z}{dx^2}$$

at all points.\* Thus we obtain the equation

$$\frac{d^2 z}{dx^2} = - \frac{Q}{EI} (a - x). \quad . \quad . \quad . \quad . \quad . \quad . \quad (63)$$

\* The negative sign comes from the fact that the curve is concave toward the negative  $z$ -axis.



Integrating twice, there results

$$z = -\frac{Q}{EI} \left( \frac{ax^2}{2} - \frac{x^3}{6} \right) + c_1x + c_2. \quad \dots \quad (64)$$

Since  $z$  and also  $dz/dx$  vanish at  $x = 0$ , both constants of integration are zero, and we have for the depression of the end of the beam (at  $x = a$ ),

$$z_a = -\frac{Q}{EI} \frac{a^3}{3}. \quad \dots \quad (65)$$

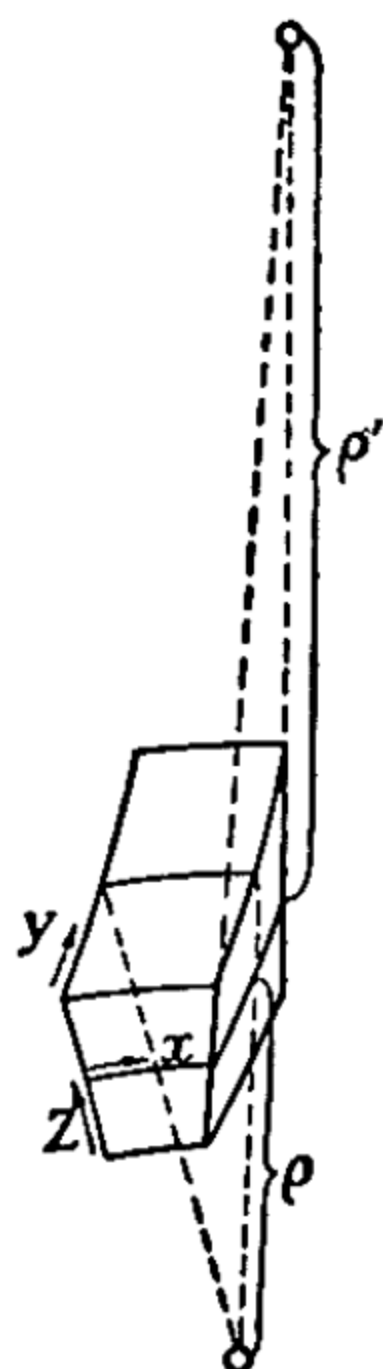


Fig. 5

The deformation of the cross-section, which was not taken into account by Bernoulli, may also be given without difficulty, provided the cross-section is initially rectangular. Since a transverse contraction takes place above the neutral surface, where the fibres are stretched, and a transverse expansion occurs below this surface, the originally vertical sides of the rectangle must converge toward the top, so that the section now has the form of a sector of a circular annulus (fig. 5). Let  $\rho'$  be the radius of curvature of the line of intersection of the neutral surface with this cross-section, and call the radius of curvature of the neutral fibre  $\rho$ , as before. Following the same method as above, we find that the relative transverse extension (in reality, contraction) of a layer which is a distance  $z$  above the neutral surface is given by

$$\frac{(\rho' - z) d\phi - \rho' d\phi}{dy} = -z \frac{d\phi}{dy} = -\frac{z}{\rho'}$$

Now the longitudinal dilatation at the same point was found to be  $+z/\rho$ , by (58). According to the definition of Poisson's Ratio, we then have

$$\sigma = -\frac{\rho}{\rho'} \quad \text{or} \quad \rho' = -\frac{\rho}{\sigma}. \quad \dots \quad (66)$$

Thus the originally plane layer at height  $z$  is bent into the form of an anticlastic surface whose two radii of curvature are in the ratio  $\sigma$ . This fact is the basis of an elegant and simple method of determining  $\sigma$  (see *Ex. 56*).

*Ex. 55.* Calculate the deflection of a tube of length  $a$ , radii  $r_i$  and  $r_o$ , supported at the ends and carrying a concentrated load at the mid-point. *HINT:* In order to reduce this to the above case, imagine the middle of the tube clamped fast, and an upward force equal to half the load applied at each end.

*Ex. 56.* A plate of rectangular section, supported along its central line, is bent by loads applied at the ends. A plane parallel plate of glass is laid upon the bent plate at the centre, and the system is illuminated from above by monochromatic

light. The interference fringes (p. 410), appearing at the thin layer between the plates, furnish a contour map of the surface resulting from the bending. By approximate representation of the surface in the neighbourhood of the point of contact, show that the contour lines are hyperbolas, the angle between the asymptotes being related to  $\sigma$  in a simple way. **HINT:** Observe that for the point of contact

$$\frac{\partial^2 z}{\partial x^2} = \frac{1}{\rho}, \quad \frac{\partial^2 z}{\partial y^2} = \frac{1}{\rho'}, \quad \text{and} \quad \rho' = -\frac{\rho}{\sigma},$$

and that  $\rho$  and  $\rho'$  may be considered constant in the neighbourhood of this point.

*Ex. 57.* A wire of length  $l$  and radius of cross-section  $a$  is clamped at one end, the other end being twisted through an angle  $\phi$ . Compute the resulting moment of torsion, which will be proportional to  $\phi$ , by imagining the wire divided into thin, concentric hollow cylinders, and these in turn subdivided into narrow rectangular elements that suffer pure shear when the wire is twisted.

## 7. Waves in Unbounded Elastic Media (Seismic Waves). Longitudinal Waves in Bars.

The fundamental equation (52) (p. 174) holds for elastokinetics as well as for elastostatics. We wish to apply it to the kinetic case and study the disturbances arising in an unbounded medium subjected to no external body forces. In this connexion, the inertial forces given by d'Alembert's principle are

$$-\rho \frac{d^2 \mathbf{s}}{dt^2}$$

per unit volume,  $\rho$  being the density. If all other body forces are absent, equation (52) becomes

$$\rho \frac{d^2 \mathbf{s}}{dt^2} = \frac{E}{2(1+\sigma)} \left\{ \Delta \mathbf{s} + \frac{1}{1-2\sigma} \text{grad div } \mathbf{s} \right\}. \quad (67)$$

From among the many integrals of this second-order partial differential equation, we select in the first instance the simple special case where the displacement vector  $\mathbf{s}$  depends only upon one co-ordinate—say  $x$ —and thus derivatives with respect to  $y$  and  $z$  vanish. In co-ordinate form, we then have

$$\left. \begin{aligned} \rho \frac{\partial^2 u}{\partial t^2} &= \frac{E}{2(1+\sigma)} \left\{ \frac{\partial^2 u}{\partial x^2} + \frac{1}{1-2\sigma} \frac{\partial^2 u}{\partial x^2} \right\} = \frac{E(1-\sigma)}{(1+\sigma)(1-2\sigma)} \frac{\partial^2 u}{\partial x^2} \\ \rho \frac{\partial^2 v}{\partial t^2} &= \frac{E}{2(1+\sigma)} \frac{\partial^2 v}{\partial x^2} \\ \rho \frac{\partial^2 w}{\partial t^2} &= \frac{E}{2(1+\sigma)} \frac{\partial^2 w}{\partial x^2} \end{aligned} \right\}. \quad (68)$$

These are, however, equations of plane waves propagated in the direction of the  $x$ -axis, as comparison with the wave equation (47) (p. 62) shows. The coefficient of  $\partial^2 u / \partial t^2$  gives the square of the reciprocal of the wave (phase) velocity, if the coefficient of  $\partial^2 u / \partial x^2$  is reduced to unity. It is evident from (68) that the wave velocity for a displacement along the  $x$ -axis—i.e. in the direction of propagation—differs from the velocity for displacements normal to this direction. Waves of the first kind are called longitudinal—those of the second kind, transverse. The wave velocities are

$$c_{\text{long}} = \sqrt{\frac{E(1-\sigma)}{\rho(1+\sigma)(1-2\sigma)}}, \quad \dots \dots \dots (69)$$

$$c_{\text{trans}} = \sqrt{\frac{E}{2\rho(1+\sigma)}} \cdot \dots \dots \dots (70)$$

There is another less formal way of arriving at the classification into longitudinal and transverse waves, which has more direct physical significance. If we replace  $\Delta s$  by  $\text{grad div } s - \text{curl curl } s$ , according to the familiar rule of vector analysis (p. 40), then (67) becomes

$$\rho \frac{\partial^2 s}{\partial t^2} = \frac{E}{2(1+\sigma)} \left\{ \frac{2(1-\sigma)}{1-2\sigma} \text{grad div } s - \text{curl curl } s \right\}. \quad \dots \dots (67')$$

We now divide  $s$  into two parts,  $s_1 + s_2$ , which are to satisfy the conditions:

$$\left. \begin{array}{ll} \text{curl } s_1 = 0, & \text{div } s_1 \neq 0 \\ \text{div } s_2 = 0, & \text{curl } s_2 \neq 0 \end{array} \right\} \cdot \dots \dots \dots (71)$$

This means that  $s_1$  represents a compressional wave, since, by equation (20) (p. 166), the volume dilatation is measured by

$$\text{div } s = \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z}.$$

At the same time, the individual volume elements experience no rotation, since  $\text{curl } s_1 = 0$ . On account of the postulated independence of  $y$  and  $z$ , this part has only an  $x$ -component:

$$\rho \frac{\partial^2 u_1}{\partial t^2} = \frac{E(1-\sigma)}{(1+\sigma)(1-2\sigma)} \frac{\partial^2 u_1}{\partial x^2}. \quad \dots \dots (72)$$

Comparison with (68) shows that the compressional wave is identical with the longitudinal wave. On the other hand, compression is absent for  $s_2$ , since  $\text{div } s_2 = 0$ . As a consequence of the vanishing of the divergence, and of the vanishing of  $\partial v_2 / \partial y$  and  $\partial w_2 / \partial z$  separately,  $\partial u_2 / \partial x$  also vanishes. The non-compressional wave of shear, as we may call this part, has therefore no component in the direction of propagation, and is thus transverse. The actual proof of the identity of this wave with the transverse waves of (68) is obtained by writing the equations for the  $y$ - and  $z$ -components of (67'), which yield

$$\rho \frac{\partial^2 v_2}{\partial t^2} = \frac{E}{2(1+\sigma)} \frac{\partial^2 v_2}{\partial x^2} \quad \text{and} \quad \rho \frac{\partial^2 w_2}{\partial t^2} = \frac{E}{2(1+\sigma)} \frac{\partial^2 w_2}{\partial x^2}. \quad \dots (73)$$

These are the equations (68) obtained above for the transverse waves.



The ratio of the velocities of propagation of compressional and shearing waves is

$$\frac{c_l}{c_t} = \sqrt{\frac{2(1-\sigma)}{1-2\sigma}}. \quad \cdot \cdot \cdot \cdot \cdot \cdot (74)$$

The value  $\sigma = 1/4$ , which holds for many substances, makes this ratio equal to  $\sqrt{3}$ . In earthquakes, both types of wave are observed; observation of the ratio of the velocities permits the calculation of  $\sigma$  for that part of the interior of the earth in which the waves travel. The value thus obtained is 0.29.

Apart from these waves in the body of the substance, there are longitudinal as well as transverse *surface waves* which follow the surface of separation of two media having different elastic properties. These waves correspond to the electrical surface waves treated on p. 345, and are called *Rayleigh waves*, after their discoverer.

The propagation of waves in an unlimited medium is to be distinguished from propagation along a bar whose transverse dimensions are small compared to its length. If we imagine the unbounded medium divided into such bars, whose axes are along the direction of propagation, we see that the boundary conditions are different for the single isolated bar and for a bar imagined cut from an unlimited medium. In the former case the transverse contraction can take place without hindrance, while the surface of the bar in the second case is elastically bound to the surroundings, and so a transverse contraction causes counter stresses. For this reason it is much easier to treat longitudinal waves in thin bars. We need not go back to the general equations, but can obtain the differential equation in the following simple way. Consider an element of the bar included between two cross-sections distant  $dx$  apart. Call the extension on the left side of this element

$$\left(\frac{\partial u}{\partial x}\right)_x,$$

and that on the right side

$$\left(\frac{\partial u}{\partial x}\right)_{x+dx}.$$

Then, if the area of the section is  $A$ , and if Young's modulus is  $E$ , the resultant force acting on the element is

$$F_x = AE \left\{ \left(\frac{\partial u}{\partial x}\right)_x + \left(\frac{\partial^2 u}{\partial x^2}\right)_x dx - \left(\frac{\partial u}{\partial x}\right)_x \right\}, \quad \cdot (75)$$

and Newton's Second Law gives

$$A\rho \frac{\partial^2 u}{\partial t^2} dx = AE \frac{\partial^2 u}{\partial x^2} dx \quad \text{or} \quad \frac{\partial^2 u}{\partial x^2} = \frac{\rho}{E} \frac{\partial^2 u}{\partial t^2}. \quad (76)$$

The velocity of propagation of longitudinal (compressional) waves in a bar is thus

$$c = \sqrt{\frac{E}{\rho}} \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad (77)$$

### 7. Transverse Vibration of Stretched Strings and Membranes.

In many cases it proves expedient to start, not with the general equations—modifying them to fit the special problem in hand—but to proceed as we did in treating longitudinal waves in bars; viz. to derive the equations immediately from the fundamental principles of the theory of elasticity, thereby taking advantage of the simplifying conditions of the problem. The transverse vibrations of stretched strings and membranes form a case in point. These vibrations are not only of great physical importance, in acoustics, but their discussion leads to mathematical operations applicable in the most diverse branches of physics, especially in atomic physics.

By a string we mean an elastic body whose cross-section is so small compared to its length, that it offers no appreciable resistance

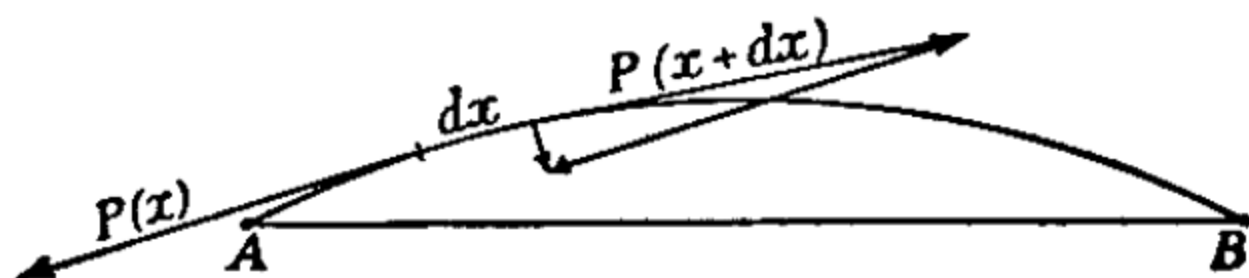


Fig. 6

to bending. The fact that the stiffness of an elastic bar diminishes as the cross-section decreases follows from the results of § 6 (p. 176). As the cross-section is made smaller and smaller, we finally have only the fibres in the immediate neighbourhood of the central filament, and there are no longer any appreciable stresses opposing the bending. How is it possible, then, for a string to execute vibrations? The state of affairs is completely altered by the fact that the string is stretched between two fixed end points. Even in the equilibrium position, there exists a stress  $P$  normal to the uniform cross-sections  $A$ , i.e. the tension  $P$  is always in the direction of the tangent to the central line of the string. Because of the complete flexibility of the string, transverse forces do not exist. If the string is now drawn aside from its position of rest, the resulting increase in length causes an increase in tension. But even this increase in tension is not the true cause of the transverse vibrations; we shall see, below, that it is a small quantity of higher order than the true cause of the vibrations. The latter is much more direct: in the strained state, the tangents to the string, drawn at the ends of an element  $dx$ , have slightly different directions. As a consequence, there is a resultant force in the direction normal to the rest position of the string (see fig. 6). If we consider only vibrations in one plane, and take the plane of vibration to be the  $xy$ -plane,

then the  $y$ -component of the displacement, which was previously denoted by  $v$ , is simply  $y$ , since the rest position of the string is the  $x$ -axis. In the displaced position, a force  $APt(x)$  acts at the left side of an element  $dx$ , if we call the unit tangent vector  $t$ . Similarly, a force  $APt(x+dx)$  acts at the right side. If  $\alpha$  is the inclination to the horizontal, the components of the resultant are

$$\left. \begin{aligned} F_x &= AP \cos(\alpha + d\alpha) - AP \cos \alpha, \\ F_y &= AP \sin(\alpha + d\alpha) - AP \sin \alpha. \end{aligned} \right\} \quad \cdot \quad \cdot \quad (78)$$

We limit the displacements to small values, so that we may set

$$\cos \alpha = 1 - \frac{\alpha^2}{2}, \quad \sin \alpha = \alpha = \tan \alpha = \frac{dy}{dx}.$$

Also, the change in the tension  $P$  is small compared with the value of the tension itself. Neglecting all quantities of second order, the resultant in the transverse direction becomes

$$F_y = AP \left\{ \left( \frac{dy}{dx} \right)_{x+dx} - \left( \frac{dy}{dx} \right)_x \right\} = AP \frac{d^2y}{dx^2} dx. \quad \cdot \quad (79)$$

(In the longitudinal direction the change in  $P$  would need to be taken into account, since only terms of second order appear here.)

Using Newton's Second Law, the equation of motion of transverse vibrations of the element becomes

$$\begin{aligned} \rho A \frac{\partial^2 y}{\partial t^2} dx &= AP \frac{\partial^2 y}{\partial x^2} dx, \\ \frac{\partial^2 y}{\partial x^2} &= \frac{\rho}{P} \frac{\partial^2 y}{\partial t^2}. \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad (80) \end{aligned}$$

This equation has the familiar form of the wave equation (cf. p. 62). It is to be noted that the tension enters here, but that the modulus of elasticity of the string does not appear.

The solution

$$y_1 = Ce^{i\omega(t-x\sqrt{\rho/P})},$$

corresponding to a wave travelling in the direction of the positive  $x$ -axis, satisfies the differential equation, but not the boundary conditions, viz.  $y$  vanishes at every instant for  $x = 0$  and for  $x = l$ , where  $l$  is the length of the string. Similarly,

$$y_2 = Ce^{i\omega(t+x\sqrt{\rho/P})}$$

is also a solution of the differential equation, and corresponds to a wave moving in the opposite direction, but it, too, fails to satisfy the end conditions. On account of the linearity of the differential



equation, the difference of these two particular solutions is also a solution. Moreover, we can give at least certain particular values of  $\omega$  for which the boundary conditions are satisfied. The difference is

$$y = y_2 - y_1 = Ce^{i\omega t} \{e^{i\omega x \sqrt{\rho/P}} - e^{-i\omega x \sqrt{\rho/P}}\} = 2iCe^{i\omega t} \sin\left(\omega x \sqrt{\frac{\rho}{P}}\right). \quad (81)$$

It is evident that the condition that  $y$  vanish for  $x = 0$  is satisfied for every value of  $t$ . On the other hand,  $y = 0$  for  $x = l$  only when  $\omega l \sqrt{\rho/P}$  is an integral multiple of  $\pi$ . If  $n$  is any positive integer, there is therefore a corresponding frequency  $\nu_n$ , given by

$$\nu_n = \frac{n}{2l} \sqrt{\frac{P}{\rho}}. \quad \cdot \cdot \cdot \cdot \cdot \cdot \quad (82)$$

We may thus consider the vibration of the string to be a standing wave resulting from superposition of two oppositely directed waves of the same frequency  $\nu_n$ . Application of the boundary conditions no longer permits all arbitrary frequencies, but only a discrete set. This set, it is true, contains infinitely many values, since  $n$  ranges from 1 to  $\infty$ .

This result may be obtained in a somewhat different way. On account of its fundamental importance, this method also will be explained here. If, as usual, we assume equation (80) to have a solution of the form

$$y = w(x) e^{i\omega t}, \quad \cdot \cdot \cdot \cdot \cdot \cdot \quad (83)$$

then we have, after removing the time factor  $e^{i\omega t}$ , the following differential equation in  $x$  alone:

$$\frac{d^2 w}{dx^2} + \frac{\omega^2 \rho}{P} w = 0. \quad \cdot \cdot \cdot \cdot \cdot \cdot \quad (84)$$

If we put 
$$\omega \sqrt{\frac{\rho}{P}} = \lambda, \quad \cdot \cdot \cdot \cdot \cdot \cdot \quad (85)$$

this becomes 
$$\frac{d^2 w}{dx^2} + \lambda^2 w = 0. \quad \cdot \cdot \cdot \cdot \cdot \cdot \quad (86)$$

This equation has a solution consistent with the boundary conditions only for certain values of the parameter  $\lambda^2$ , which are called *characteristic values*, or *proper values*. Taking, as a solution,

$$w = \sin \lambda x, \quad \cdot \cdot \cdot \cdot \cdot \cdot \quad (87)$$

we see that  $\lambda$  must be an integral multiple of  $\pi/l$ , in order that  $y$  vanish for  $x = l$ , i.e.

$$\lambda_n = 2\pi\nu_n \sqrt{\frac{\rho}{P}} = \frac{n\pi}{l}, \quad \cdot \cdot \cdot \cdot \cdot \cdot \quad (88)$$

and thus, as above, 
$$\nu_n = \frac{n}{2l} \sqrt{\frac{\bar{P}}{\rho}}. \quad . \quad . \quad . \quad . \quad . \quad (89)$$

The functions which correspond to the characteristic values, and which are solutions of the differential equation, and which—in addition—satisfy the boundary conditions, are called the *characteristic functions*, or *proper functions*, of the differential equation.\* In the present case, the characteristic functions are the functions

$$\sin \frac{n\pi}{l} x.$$

Since the real as well as the imaginary parts of equation (83) must satisfy the differential equation separately, a solution which satisfies the boundary conditions is given by

$$y = a_n \cos 2\pi\nu_n t \sin \frac{n\pi x}{l} \quad . \quad . \quad . \quad . \quad . \quad (90)$$

as well as by 
$$y = b_n \sin 2\pi\nu_n t \sin \frac{n\pi x}{l} \quad . \quad . \quad . \quad . \quad . \quad (91)$$

But the series

$$y = \sum_{n=1}^{n=\infty} a_n \cos 2\pi\nu_n t \sin \frac{n\pi x}{l} + \sum_{n=1}^{n=\infty} b_n \sin 2\pi\nu_n t \sin \frac{n\pi x}{l} \quad (92)$$

is also a solution, since a sum of particular solutions also satisfies the equation. In the case of an infinite sum, it must also be provided that the series converges. The coefficients  $a_n$  and  $b_n$  are determined by giving the initial displacement  $y_0$  and initial velocity  $\dot{y}_0$  as functions of  $x$ : i.e. apart from the spatial boundary conditions to be fulfilled at every instant, certain initial conditions must also be satisfied. Now

$$y_0(x) = \sum_{n=1}^{n=\infty} a_n \sin \frac{n\pi x}{l}, \quad . \quad . \quad . \quad . \quad . \quad (93)$$

$$\dot{y}_0(x) = \sum_{n=1}^{n=\infty} 2\pi\nu_n b_n \sin \frac{n\pi x}{l}. \quad . \quad . \quad . \quad . \quad (94)$$

The determination of the coefficients follows the method developed for those of the Fourier series (p. 54). The right member of (93) is simply a Fourier series which is to represent the given function in the interval 0 to  $l$ . To determine the coefficient  $a_m$ , for example, we multiply both sides of the first equation by

$$\sin \frac{m\pi x}{l}$$

\* The German expressions for "characteristic value" and "characteristic function" are *Eigenwert* and *Eigenfunktion*.

and integrate from 0 to  $l$ . All terms on the right vanish, with the exception of that containing

$$\frac{m\pi x}{l},$$

which, by p. 55, gives the following value for  $a_m$ :

$$a_m = \frac{2}{l} \int_0^l y_0(x) \sin \frac{m\pi x}{l} dx. \quad \dots \quad (95)$$

In similar manner, the second equation gives

$$b_m = \frac{1}{\pi \nu_m l} \int_0^l \dot{y}_0(x) \sin \frac{m\pi x}{l} dx. \quad \dots \quad (96)$$

The vibrations corresponding to the various frequencies are of different strength, depending upon the initial conditions. The vibrations of the string are transmitted to the ear by the intervening air, and—in so far as they lie within the audible range—are perceived as a tone whose pitch is given by the fundamental frequency  $\nu_1$ , and whose quality, or timbre, is determined by the amplitudes of the various harmonics, with frequencies  $\nu_n$  which are multiples of  $\nu_1$ . Thus equations (95) and (96) are the mathematical expressions of the fact that a given string tone sounds different, according as it is produced by plucking or by striking the string. As an example of the determination of the coefficients, let us choose the struck string; this corresponds, say, to a pianoforte string. For  $t = 0$ , let  $y$  be zero everywhere. Let  $\dot{y}_0$  be zero at all points except between  $(l-a)/2$  and  $(l+a)/2$ , where it is to have the value  $v_0$ . Thus  $a$  is the width of the hammer. Then  $a_m = 0$  and

$$\begin{aligned} b_m &= \frac{1}{\pi \nu_m l} \int_{(l-a)/2}^{(l+a)/2} v_0 \sin \frac{m\pi x}{l} dx \\ &= \frac{v_0}{m\pi^2 \nu_m} \left\{ \cos \frac{m\pi}{l} \left( \frac{l-a}{2} \right) - \cos \frac{m\pi}{l} \left( \frac{l+a}{2} \right) \right\}. \end{aligned} \quad (97)$$

If the string is not struck in the middle, as assumed here, the coefficients will have other values.

It is not difficult to treat the two-dimensional analogue—the vibrating membrane—in the same way. Here, again, let the membrane be so thin that it possesses no appreciable stiffness. As in the case of the string, it is only because the membrane is fastened at the edges that restoring forces arise. Assume the bounding curve to be plane, its plane being the  $xy$ -plane of our system of co-ordinates. The constant tension  $P$  acts upon every element of area in a section normal to the plane of the membrane. Consider an element  $dx dy$  in the dis-



placed position of the membrane. Then the tensions parallel to the  $yz$ -plane have a resultant

$$F_{1z} = DP \frac{\partial^2 z}{\partial y^2} dx dy,$$

where  $D$  is the thickness of the membrane. The tensions in a direction normal to this have a resultant amounting to

$$F_{2z} = DP \frac{\partial^2 z}{\partial x^2} dx dy.$$

Altogether, we have a total force acting in the direction of the  $z$ -axis of amount

$$F_z = DP dx dy \left( \frac{\partial^2 z}{\partial y^2} + \frac{\partial^2 z}{\partial x^2} \right).$$

Applying Newton's Second Law, the equation of motion for the element under consideration becomes

$$\rho D dx dy \frac{\partial^2 z}{\partial t^2} = DP dx dy \left( \frac{\partial^2 z}{\partial y^2} + \frac{\partial^2 z}{\partial x^2} \right)$$

$$\text{or} \quad \frac{\partial^2 z}{\partial x^2} + \frac{\partial^2 z}{\partial y^2} = \frac{\rho}{P} \frac{\partial^2 z}{\partial t^2}. \quad \cdot \cdot \cdot \cdot \cdot (98)$$

The spatial boundary condition to be satisfied in addition to the differential equation is that  $z = 0$  along the bounding curve at all times. Generally, at the boundary, we could prescribe definite *temporally constant* values of  $z$  or of  $\partial z / \partial x$ , but not of both, as one might conclude from the degree of the differential equation. It is easy to see that the problem would be over-determined by specifying *both*  $z$  and  $\partial z / \partial x$ . Imagine the integration carried out numerically or graphically, piece by piece, in a strip parallel to the  $y$ -axis. We begin with a surface element whose position and orientation are given, and join the next element to this, the orientation being determined by the differential equation, which gives the increase of  $\partial z / \partial x$ . If we proceed in this way, we shall not necessarily encounter the boundary a second time; hence only the position *or* the orientation may be specified for the bounding curve.

We again assume that the solution is of the form

$$z = w(x, y)e^{i\omega t}, \quad \cdot \cdot \cdot \cdot \cdot (99)$$

and obtain, just as in the case of the string,

$$\frac{\partial^2 w}{\partial x^2} + \frac{\partial^2 w}{\partial y^2} + \lambda^2 w = 0, \quad \lambda = \omega \sqrt{\frac{\rho}{P}}. \quad \cdot \cdot (100)$$

In general, this equation, too, has a solution which satisfies the boundary conditions only for a discrete sequence of characteristic values  $\lambda_n^2$ .

The nature of the corresponding characteristic functions depends upon the boundary curve. They may be complicated higher transcendental functions—e.g. Bessel functions, in the case of a circular boundary. Nevertheless, for given initial conditions, the coefficients of a series composed of characteristic functions may be determined as for the Fourier series. Thus it can be shown that all the characteristic functions of the equation  $\Delta w + \lambda^2 w = 0$ , which vanish on the boundary, have the property of orthogonality; i.e. in analogy to the vanishing of the integrals

$$\int_0^l \sin \frac{m\pi x}{l} \sin \frac{n\pi x}{l} dx,$$

we have in general

$$\int f_m f_n d\tau = 0, \quad . \quad . \quad . \quad . \quad . \quad (101)$$

the integral being extended over the entire range of  $w$ , i.e. in our case, over the area of the membrane. We give at once the proof for the three-dimensional case: If  $f_m$  and  $f_n$  are two different characteristic functions, then, by Gauss's Theorem,

$$\oint (f_m \text{ grad } f_n - f_n \text{ grad } f_m) d\mathbf{S} = \int (f_m \Delta f_n - f_n \Delta f_m) d\tau, \quad (102)$$

which, by (100), is equal to

$$\int (\lambda_m^2 - \lambda_n^2) f_m f_n d\tau. \quad . \quad . \quad . \quad . \quad . \quad (103)$$

The surface integral is to be extended over the boundary of the region, the volume integral over the region itself. But  $f_m$  and  $f_n$  are always zero on the boundary, and so the left member of (102) is zero. Since  $\lambda_m$  and  $\lambda_n$  are in general different, we have

$$\int f_m f_n d\tau = 0. \quad . \quad . \quad . \quad . \quad . \quad (104)$$

If we now write an arbitrary point function  $g(x, y, z)$  as the sum of characteristic or proper functions  $f_i$ ,

$$g(x, y, z) = a_1 f_1 + a_2 f_2 + \dots + a_m f_m + \dots, \quad (105)$$

we may proceed to determine the coefficients exactly as with Fourier's series. If we multiply both sides by  $f_m$  and integrate over the entire bounded region (in the case of the membrane, over the entire area), the orthogonality makes all terms of the series on the right, except

$$\int a_m f_m^2 d\tau,$$

vanish. The coefficients are thus

$$a_m = \frac{\int g(x, y, z) f_m d\tau}{\int f_m^2 d\tau}. \quad \cdot \cdot \cdot \cdot (106)$$

The characteristic functions are determined only to a constant factor. This multiplier is usually so chosen that the integral

$$\int f_m^2 d\tau$$

has the value 1. In this case the characteristic or proper functions are said to be normalized to unity. Consult the additional discussion on p. 703.

*Ex. 58.* Give the solution of the differential equation for a rectangular membrane of dimensions  $a$  and  $b$ . [Try a solution of the form  $w(x, y) = p(x)q(y)$ .]



## CHAPTER IX

### THE MECHANICS OF LIQUIDS AND GASES (HYDRO- AND AERO-MECHANICS)

#### 1. Equilibrium of Fluid Bodies (Hydrostatics).

The liquid and gaseous states—especially the latter—are characterized by the fact that no resistance is offered to a change in shape which is unaccompanied by a change in volume. While liquids have extremely small compressibilities (i.e. the forces opposing a change in volume are large), those of gases are very great. Moreover, gases tend to expand, filling the entire available volume. If we apply to fluids the theory developed in the last chapter for deformable bodies, then the absence of rigidity of form is expressed by the fact that the shearing stresses  $P_{12}$ ,  $P_{23}$ ,  $P_{31}$  are zero for an ideal non-viscous fluid. (It is to be remembered that the term “fluid” includes both the liquid and gaseous states.) Actual fluids, it is true, are subject to tangential forces when in motion. If two layers slide over each other with a certain relative speed, they exert upon each other tangential forces proportional to the relative speed. Thus these forces vanish for a state of rest, so that it is not necessary to distinguish between viscous and non-viscous fluids in hydrostatics. The vanishing of the tangential (or shearing stresses)  $P_{12}$ ,  $P_{23}$ ,  $P_{31}$  implies the vanishing of the terms in  $xy$ ,  $yz$  and  $zx$  in the equation of the tensor ellipsoid, and this is true for every co-ordinate system. This means that every co-ordinate system is a system of principal axes, which is possible only if the ellipsoid is a sphere. Hence, in this case, all the normal stresses  $P_{11}$ ,  $P_{22}$ ,  $P_{33}$  must be equal. We set them equal to  $-p$ .\* Thus a normal *pressure*  $p$ , of equal strength in all directions, acts upon any element of volume of a fluid. As a consequence, the general conditions of equilibrium of a deformable body become greatly simplified. The system of equations (35) (p. 170) becomes

$$g = \text{grad } p. \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

In hydromechanics it is customary to refer the body forces to unit

\* Since the normal stresses in the theory of elasticity are reckoned positive when in the direction of the *outward* drawn normal to a closed surface, we must give the *inward* pressure  $p$  the negative sign.

mass rather than to unit volume. If  $\rho$  is the density of the fluid, and  $\mathbf{G}$  is the body force acting upon one gramme of fluid, then the body force on unit volume is  $\rho\mathbf{G}$ , and the fundamental equation of hydrostatics becomes

$$\mathbf{G} = \frac{1}{\rho} \text{grad } p. \quad . \quad . \quad . \quad . \quad . \quad (1a)$$

Just as it was necessary to introduce a relationship between the stress and strain tensors in elastomechanics, so a connexion between density and pressure must be supplied in hydrostatics. In general, the compressibility of liquids may be neglected, and so  $\rho$  may be considered constant; for gases, the required connexion is given by Boyle's Law (see below).

Of the many special problems of hydrostatics and aerostatics we select two—the pressure gradient in the atmosphere and the form of the surface of a rotating liquid. For the first problem, we take a  $z$ -axis vertically upward, and the  $x$ - and  $y$ -axes in the horizontal plane, which is the surface of the earth, since we limit ourselves to a small portion of the surface. If the density at a height  $z$  is  $\rho(z)$ , then the force acting upon unit mass of air is  $-g\mathbf{k}$ . Since  $x$  and  $y$  do not enter,  $\text{grad } p$  reduces to  $(dp/dz)\mathbf{k}$ , and we have the equation

$$\frac{dp}{dz} = -\rho g. \quad . \quad . \quad . \quad . \quad . \quad (2)$$

Now for gases at constant temperature, the density is proportional to the pressure, by Boyle's Law; that is,

$$\frac{\rho}{\rho_0} = \frac{p}{p_0}. \quad . \quad . \quad . \quad . \quad . \quad (3)$$

Using this relationship, equation (2) becomes

$$\frac{dp}{p} = -\frac{\rho_0}{p_0} g dz.$$

Integrating between  $z = 0$  (at the earth's surface) and  $z = z$ ,

$$\log \frac{p}{p_0} = -\frac{\rho_0}{p_0} gz$$

or 
$$p = p_0 e^{-(\rho_0/p_0)gz}. \quad . \quad . \quad . \quad . \quad . \quad (4)$$

This formula is due to Halley. It is to be remembered that Halley's formula is true only provided the temperature is constant at all levels. Nevertheless, it furnishes a fairly good representation of observed pressures at various altitudes.

The integration of the hydrostatic equation is particularly simple

if the density may be considered constant and if the external forces are derivable from a potential  $U$ . In this case

$$\mathbf{G} = -\text{grad } U, \quad . . . . . (5)$$

and (1a) becomes

$$\text{grad } U + \frac{1}{\rho} \text{grad } p = 0.$$

If  $\rho$  is constant, this equation may be integrated at once, giving

$$U + \frac{p}{\rho} = \text{const.} \quad . . . . . (6)$$

This equation states that, under the assumptions made, the surfaces of equal pressure are equipotential surfaces of the acting forces. Now the free surface of a liquid is always a surface of constant pressure, the pressure being that of the atmosphere. Hence *the free surface is an equipotential surface*. For a fluid at rest and subject only to gravity, this surface of equal potential is a sphere with centre at the earth's centre and radius equal to the distance of the surface from this point.

We next investigate the form of the surface of a liquid contained in a vertical cylindrical vessel and rotating, along with the container, about the axis of the cylinder with angular speed  $\omega$ . This is only apparently a dynamical phenomenon. After a brief period of adjustment there results a state where the liquid rotates as a whole, along with the vessel, and relative movement of the particles no longer occurs. Thus an observer rotating with the system notes a state of equilibrium. Apart from gravity, the only force acting upon the individual volume elements is the centrifugal force, which we may treat as an externally applied force also. Taking the positive  $z$ -axis upward along the axis of the cylinder, the potential of gravity becomes  $gz$ . The centrifugal force is radial, and so is normal to the cylinders

$$\sqrt{x^2 + y^2} = r = \text{const.}$$

The force vector is directed outward and has the magnitude  $\omega^2 r$ , and hence is the negative of the gradient of a function

$$U = -\frac{\omega^2 r^2}{2}.$$

Then, according to equation (6), the surfaces of equal pressure, including the free surface, are given by

$$z - \frac{\omega^2 r^2}{2g} = \text{const.}, \quad . . . . . (7)$$

i.e. they are paraboloids of rotation whose meridian section has the equation

$$z - \frac{\omega^2}{2g} x^2 = \text{const.} \quad . . . . . (7')$$



If the origin of co-ordinates is placed at the lowest point (vertex) of this curve, the equation of the surface is

$$z = \frac{\omega^2}{2g} x^2.$$

The vertex of the parabola is, of course, below the plane of the non-rotating liquid, while the liquid at the walls is higher than this plane.

It may be mentioned that the phenomenon just discussed has been used to produce a concave parabolic mirror for optical purposes by rotating a vessel of mercury. The method, however, was not very successful.

*Ex. 59.* Using the hydrostatic equations, derive *Archimedes' Principle*, which states that a body immersed in a fluid is buoyed up by a force equal to the weight of the displaced fluid. (Remember that the upward force is the resultant of the pressure forces.)

## 2. The Fundamental Hydrodynamical Equations.

If a fluid is not in a state of rest (equilibrium), we say that a condition of flow prevails. This state of flow is completely defined only if the flow velocity is known at every point at each instant. In practice, however, we deal almost wholly with problems of stationary states, i.e. with cases in which the flow velocity vector at a given point does not change with time. If, for example, we inquire about the lift experienced by an aeroplane flying with uniform velocity, we may consider the aircraft stationary and the air blowing past it from infinity with uniform velocity. This is actually the arrangement used in wind tunnel experiments. The distribution of flow about the aeroplane is then steady. In the non-stationary state, two kinds of curve are to be distinguished—the *lines of motion*, or paths described by single particles of the fluid in the course of time, and the *stream lines*, which are defined as follows: At a given instant let the flow velocity vectors at each point be drawn in the form of small arrows. These arrows may be put together to form curves whose tangents at each point have the same direction as the arrows. In the non-stationary state, the picture of the stream lines changes from instant to instant. Consequently, a particle which moves from  $P$  to  $P'$  in a time  $dt$  has a velocity at  $P'$  different from that given by the stream-line picture corresponding to the time when the particle was still at  $P$ . In this way we see that *the lines of motion and the stream lines coincide only in the stationary state*.

In hydromechanics, just as in elastomechanics, we perform the transition from statics to kinetics by introducing the inertial force for unit volume,

$$-\rho \frac{d^2 \mathbf{r}}{dt^2},$$

by means of d'Alembert's principle, thus referring the kinetic problem back to statics. There will be equilibrium between this force, the body forces and the pressure forces when, according to equation (1a) (p. 192),

$$\rho \mathbf{G} - \rho \frac{d^2 \mathbf{r}}{dt^2} = \text{grad } p,$$

or 
$$\frac{d^2 \mathbf{r}}{dt^2} = \mathbf{G} - \frac{1}{\rho} \text{grad } p. \quad . . . . . (8)$$

In anticipation of the most important special problem of stationary flow we resolve the total rate of increase of velocity per unit time, i.e. the vector  $d^2 \mathbf{r}/dt^2 = d\mathbf{v}/dt$ , into a purely temporal part and a purely spatial part. The latter arises from the fact that, in the time  $dt$ , the volume element moves to a place  $\mathbf{r} + d\mathbf{r}$ , where the velocity is  $\mathbf{v} + d\mathbf{v} = \mathbf{v} + d\mathbf{r} \text{grad } \mathbf{v}$ . Reduced to unit time, this part of the acceleration is

$$\frac{d_r \mathbf{v}}{dt} = \frac{d\mathbf{r}}{dt} \text{grad} \cdot \mathbf{v} = \mathbf{v} \text{grad} \cdot \mathbf{v}, \quad . . . . . (9)$$

where  $d_r$  signifies the *spatial* change.

For non-stationary motion, on the other hand, the flow configuration changes during the time  $dt$ , so that the particle at  $\mathbf{r} + d\mathbf{r}$  is in a state corresponding to the time increment  $dt$ . Neglecting terms of higher order, we may put for this purely temporal change the value obtaining at the point  $\mathbf{r}$ . Let this acceleration be denoted by  $\partial \mathbf{v}/\partial t$ . The total acceleration is equal to the sum of the purely temporal and purely spatial rates of change of velocity, and so equation (8) becomes

$$\frac{\partial \mathbf{v}}{\partial t} + \mathbf{v} \text{grad} \cdot \mathbf{v} = \mathbf{G} - \frac{1}{\rho} \text{grad } p. \quad . . . . . (10)$$

Besides this fundamental hydrodynamical equation we have another, called the equation of continuity. The mass of fluid passing *outward* through the surface of an arbitrary volume element, per second, is

$$\oint \rho \mathbf{v} d\mathbf{S}.$$

The scalar product  $\mathbf{v} d\mathbf{S}$  automatically provides that *inward* flow be reckoned negative. The quantity flowing out must be equal to the decrease, per second, in the amount of fluid within the volume element, provided we rule out the possibility of the existence of sources or sinks within the volume. That is,

$$\oint \rho \mathbf{v} d\mathbf{S} = - \frac{\partial}{\partial t} \int \rho d\tau. \quad . . . . . (11)$$

If we convert the surface integral into a volume integral, using Gauss's theorem, then, since the equation must hold for every element of volume, we have

$$\operatorname{div}(\rho \mathbf{v}) = -\frac{\partial \rho}{\partial t}. \quad . \quad . \quad . \quad . \quad . \quad (12)$$

For incompressible fluids this equation of continuity simplifies to

$$\operatorname{div} \mathbf{v} = 0. \quad . \quad . \quad . \quad . \quad . \quad (12')$$

In view of the meaning of  $\mathbf{v} \operatorname{grad} \cdot \mathbf{v}$  given on p. 32, the component form of the general hydrodynamical equations may now be written

$$\left. \begin{aligned} \frac{\partial v_x}{\partial t} + v_x \frac{\partial v_x}{\partial x} + v_y \frac{\partial v_x}{\partial y} + v_z \frac{\partial v_x}{\partial z} &= G_x - \frac{1}{\rho} \frac{\partial p}{\partial x} \\ \frac{\partial v_y}{\partial t} + v_x \frac{\partial v_y}{\partial x} + v_y \frac{\partial v_y}{\partial y} + v_z \frac{\partial v_y}{\partial z} &= G_y - \frac{1}{\rho} \frac{\partial p}{\partial y} \\ \frac{\partial v_z}{\partial t} + v_x \frac{\partial v_z}{\partial x} + v_y \frac{\partial v_z}{\partial y} + v_z \frac{\partial v_z}{\partial z} &= G_z - \frac{1}{\rho} \frac{\partial p}{\partial z} \end{aligned} \right\}, \quad . \quad (10a)$$

and the equation of continuity becomes

$$\frac{\partial(\rho v_x)}{\partial x} + \frac{\partial(\rho v_y)}{\partial y} + \frac{\partial(\rho v_z)}{\partial z} = -\frac{\partial \rho}{\partial t}. \quad . \quad . \quad . \quad (12a)$$

The co-ordinate form of the fundamental hydrodynamical equations indicates the special mathematical difficulties arising in hydrodynamics. Since products of velocity components with their derivatives occur, the equations, as compared with those which have previously occurred, are not linear, and the rule for writing the general solution as the sum of particular solutions is no longer valid. This rule, which has wide applications to linear partial differential equations occurring in physics, e.g. the electrical field equations, has been used above, in the treatment of the vibrating string (p. 186).

We limit ourselves, in what now follows, to the case of stationary flow, and so put all pure time derivatives equal to zero. We again transform the resulting equations for the stationary state and obtain a further classification of the flow patterns into *rotational* and *irrotational* flow. From the vector formula

$$[\mathbf{A} \operatorname{curl} \mathbf{B}] = [\mathbf{A} [\nabla \mathbf{B}]] = \nabla \cdot \mathbf{A}_c \mathbf{B} - \mathbf{A} \nabla \cdot \mathbf{B} = \operatorname{grad} \mathbf{A}_c \mathbf{B} - \mathbf{A} \operatorname{grad} \mathbf{B}$$

derived on p. 38, we have for  $\mathbf{B} = \mathbf{A}$ :

$$[\mathbf{A} \operatorname{curl} \mathbf{A}] = \operatorname{grad} \mathbf{A}_c \mathbf{A} - \mathbf{A} \operatorname{grad} \cdot \mathbf{A} = \frac{1}{2} \operatorname{grad} \mathbf{A}^2 - \mathbf{A} \operatorname{grad} \cdot \mathbf{A}.$$



If we apply this formula to the expression  $\mathbf{v} \text{ grad } \mathbf{v}$ , we have, for stationary flow,

$$\frac{1}{2} \text{grad } v^2 - [\mathbf{v} \text{ curl } \mathbf{v}] = \mathbf{G} - \frac{1}{\rho} \text{grad } p. \quad . \quad . \quad (13)$$

The integration of the equations becomes relatively simple if we are dealing with a type of flow for which  $\text{curl } \mathbf{v}$  vanishes in the entire space. Thus, if  $\text{curl } \mathbf{v}$  is zero everywhere, then  $\mathbf{v}$  itself may be represented as the gradient of a scalar quantity  $\Phi$ , the velocity potential, and the problem is solved when the single scalar quantity  $\Phi$  has been found as a function of the co-ordinates.

*Ex. 60.* Using equations (10a), calculate the equilibrium rate of flow of a liquid from a hole in the bottom of a vessel filled to a depth  $h$ .

*Ex. 61.* In the same way, calculate the velocity of escape of a gas confined under a pressure  $p_1 - p_0$  in excess of outside pressure. Assume that Boyle's Law,  $p/\rho = \text{const.}$ , holds here, and neglect the gravity force acting upon the gas.

### 3. Irrotational Flow.

#### (a) Bernoulli's Equation

If the body force per gramme is derivable from a potential, which is almost always the case, we have  $\mathbf{G} = -\text{grad } U$ . In this instance, if we integrate equation (13) along a stream line,  $d\mathbf{s}$  and  $\mathbf{v}$  have the same direction, and so  $[\mathbf{v} \text{ curl } \mathbf{v}]d\mathbf{s}$  vanishes. Then

$$\begin{aligned} \int_{P_0}^P \frac{1}{2}(\text{grad } v^2) d\mathbf{s} - \int_{P_0}^P [\mathbf{v} \text{ curl } \mathbf{v}] d\mathbf{s} \\ = \int_{P_0}^P (-\text{grad } U) d\mathbf{s} - \int_{P_0}^P \left( \frac{1}{\rho} \text{grad } p \right) d\mathbf{s} \quad . \quad (14) \end{aligned}$$

becomes, for incompressible fluids,

$$\frac{v^2}{2} + U + \frac{p}{\rho} = \frac{v_0^2}{2} + U_0 + \frac{p_0}{\rho} = \text{const.} \quad . \quad . \quad (15)$$

The values of the constants are in general different for each stream line. However, if the flow is irrotational—i.e. if  $\text{curl } \mathbf{v}$  vanishes throughout the entire liquid—then, by equation (14), the integration may be performed along any arbitrary curve. In this case, the constants must be the same for all stream lines. Reciprocally, we may conclude that if the flow originates in an irrotational\* region, where the quantity

$$\frac{v_0^2}{2} + U_0 + \frac{p_0}{\rho}$$

\* This limitation must be made on account of the possibility of  $\text{curl } \mathbf{v}$  and  $\mathbf{v}$  being parallel.

has the same value for all stream lines, then the flow remains irrotational in the entire region. This is true, for example, for the outflow of liquid from an orifice in a vessel, since all three terms of the expression

$$\frac{v_0^2}{2} + U_0 + \frac{p_0}{\rho}$$

have the same values all over the free surface.

(b) *Calculation of irrotational (or potential) flow*

The Bernoulli equation (15) shows that irrotational states of flow are not only mathematical simplifications, but also that they must occur very frequently in nature. If, for example, air is blown past an aeroplane model in wind-tunnel experiments to determine the lift, then it is true that the air entering the canal has a given velocity and a given pressure, and that the gravitational potential—which, moreover, is negligible here—has practically the same value for all stream lines. Consequently, according to the above laws, that flow pattern which corresponds to the solution based upon the assumption of absence of rotation is the one which will form about the model.

This pure potential flow, however, leads to paradoxical results. For example, imagine an infinitely long circular cylinder immersed in a uniform, parallel flow which is normal to the axis of the cylinder. According to the above principles, the fluid will exert no force on the cylinder. The discrepancy is to be attributed to the fact that even for fluids of low viscosity the friction at the surface of the body is not negligible. According to experience, the liquid clings to the body in the immediate neighbourhood of the surface, while pure potential flow requires smooth, unhindered motion. Due to the clinging of the liquid there is a large velocity gradient in the boundary layer and  $\text{curl } \mathbf{v}$  no longer vanishes there. Although the viscosity is small, it causes the boundary layer—which in the limit is to be thought of as a surface of discontinuity (see p. 40)—to roll up and detach itself, producing individual vortices. Such surfaces of discontinuity also can form where two streams of different velocity meet, e.g. behind the trailing edge of an airfoil. The fact that the methods of potential flow are still usable, especially in the two-dimensional case, is due to the possibility of representing vortices mathematically and correcting in this way the potential flow. Such cases, however, are no longer irrotational in the strict sense of (a).\*

Let us now consider the mathematical treatment of irrotational flow. We put

$$\mathbf{v} = \text{grad } \Phi. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (16)$$

\* In many books on hydrodynamics the term "turbulent" is reserved for vortex motion in the narrower sense, while parallel flow having a velocity gradient normal to the flow, so that  $\text{curl } \mathbf{v} \neq 0$ , is referred to as "non-irrotational".

The equation of continuity for incompressible fluids then becomes

$$\operatorname{div} \mathbf{v} = \operatorname{div} \operatorname{grad} \Phi = \Delta \Phi = 0, \quad . . . \quad (17)$$

and the actual hydrodynamical equations become

$$\frac{1}{2} \operatorname{grad} v^2 = \mathbf{G} - \frac{1}{\rho} \operatorname{grad} p \quad . . . \quad (18)$$

for steady flow. Thus every function  $\Phi$  which satisfies the differential equation  $\Delta \Phi = 0$  represents the velocity potential of an irrotational fluid motion. The problem is then to find the solution which corresponds to the physical conditions of the special case under consideration. With this solution at hand, the velocity field may be obtained by taking the gradient, and the actual hydrodynamical equations (18) are necessary only for the calculation of the pressure distribution. As the simplest example of potential flow in space, we investigate the spherically symmetric solution of the equation  $\Delta \Phi = 0$ . Since  $\Phi$  depends on  $r$  alone, the equation has the familiar form (cf. p. 63)

$$\frac{d^2 \Phi}{dr^2} + \frac{2}{r} \frac{d\Phi}{dr} = 0. \quad . . . \quad (19)$$

As is at once evident, a solution is

$$\Phi = \frac{a}{r} + b. \quad . . . \quad (20)$$

By taking the gradient we obtain

$$\mathbf{v} = -\frac{a}{r^2} \mathbf{r}_0. \quad . . . \quad (21)$$

This means that the flow is directed radially outward from (or inward toward) the point  $O$ . In the former instance, a source exists at  $O$ —in the second case, a sink. The constant  $a$  is determined by the volume of fluid  $Q$  flowing through a sphere about  $O$  each second. This amounts to

$$Q = -4\pi r^2 \frac{a}{r^2} = -4\pi a. \quad . . . \quad (22)$$

Thus  $a$  is determined if  $Q$  is given. The additive constant  $b$  does not affect the flow, since it drops out when the gradient is taken. Usually, however, the potential is so normalized that it vanishes at infinity, in which case  $b = 0$ .

A problem which is rather more troublesome to solve is that of the flow pattern resulting when a spherical obstacle is placed in a



parallel stream whose potential function is  $\Phi = v_0 x$ . The special conditions in this case are: At a great distance from the sphere, the fluid motion must be the undisturbed parallel flow; and at the surface of the sphere, the normal components of the flow must vanish. The fluid, considered to be ideal and non-viscous, may move freely *along* the surface of the sphere. In this connexion, the newer hydrodynamics, which corresponds better with observation, differs from the classical theory. The more recent theory starts with the assumption of a clinging boundary layer, as indicated above (p. 198).

We shall not reproduce the solution of this problem of the sphere here, since the essential point may be seen from the two-dimensional problem of the flow past an infinitely long cylinder, viz. that the resultant force on the obstacle vanishes, which is, of course, completely at variance with experience.

If the flow is taken to be two-dimensional, i.e. if the flow pattern is the same in all planes normal to the  $z$ -axis, then a large number of solutions of the potential equation can be obtained with the aid of conformal representation (p. 69). The connexion between the two-dimensional potential problem and conformal mapping of the complex plane is as follows.

Introducing the potential function, we have

$$v_x = \frac{\partial \Phi}{\partial x}, \quad v_y = \frac{\partial \Phi}{\partial y}, \quad \dots \dots \dots (16')$$

and  $\operatorname{div} \mathbf{v} = \Delta \Phi = 0$  (equation of continuity). (17)

The condition for irrotational motion is

$$\frac{\partial v_y}{\partial x} - \frac{\partial v_x}{\partial y} = 0, \quad \dots \dots \dots (23)$$

the co-ordinate  $z$  being absent. This condition is automatically fulfilled by our assumption. If, now, we take another function  $\Psi$  and put

$$v_x = \frac{\partial \Psi}{\partial y}, \quad v_y = -\frac{\partial \Psi}{\partial x}, \quad \dots \dots \dots (24)$$

then the equation of continuity is identically satisfied. In this case, the condition that there be no rotation requires of  $\Psi$  that it satisfy the equation  $\Delta \Psi = 0$ . If, then, in the two-dimensional case, we have any function  $\Omega$  which satisfies the potential equation, we can obtain a system of stream lines from it in two ways—first, by taking the gradient; secondly, by means of the correspondence (24). If, reciprocally, a given flow is specified on one hand by the potential function  $\Phi$ , and on the other hand by the *stream function*  $\Psi$  according to equation (24), then

since the components  $v_x$  and  $v_y$  obtained in these two ways must be equal, we have

$$\left. \begin{aligned} \frac{\partial \Phi}{\partial x} &= \frac{\partial \Psi}{\partial y} \\ \frac{\partial \Phi}{\partial y} &= -\frac{\partial \Psi}{\partial x} \end{aligned} \right\} \dots \dots \dots (25)$$

But these are precisely the Cauchy-Riemann differential equations (see p. 71), satisfied by the real part  $\Phi$  and the imaginary part  $\Psi$  of a function of the complex variable  $z = x + iy$ . Since, in addition,  $\Phi$  and  $\Psi$  satisfy Laplace's equation, we may say that a flow pattern is given by every complex function

$$\Omega = \Phi + i\Psi = f(x + iy); \quad \dots \dots \dots (26)$$

the curves  $\Phi = \text{const.}$  are the equipotential lines, and their orthogonal trajectories are the stream lines. But according to the theory of conformal mapping, the curves  $\Phi = \text{const.}$  are orthogonal to the curves  $\Psi = \text{const.}$ , so that the curves  $\Psi = \text{const.}$  are really the stream lines. By interchanging the functions  $\Phi$  and  $\Psi$  it follows that a second flow pattern is given by the conformal representation if we take the curves  $\Psi = \text{const.}$  to be the equipotentials and the curves  $\Phi = \text{const.}$  to be the stream lines. If we take the derivative\*

$$w = \frac{d\Omega}{dz} = \frac{\partial \Omega}{\partial x} = \frac{\partial \Phi}{\partial x} + i \frac{\partial \Psi}{\partial x} = v_x - iv_y, \quad \dots \dots (27)$$

we see that the complex number  $w$  agrees in magnitude with the flow velocity, and is the reflection of the vector  $v$  in the  $x$ -axis. Hence  $w$  is called the *conjugate complex velocity*;  $\Omega$  is called the *complex potential*.

We now take a particular form for  $f(z)$ , which, as will be seen at once, corresponds to cylindrical symmetry, viz.

$$f(z) = \Phi + i\Psi = a \log z \quad \text{or} \quad z = e^{(\Phi + i\Psi)/a} = e^{\Phi/a} \left( \cos \frac{\Psi}{a} + i \sin \frac{\Psi}{a} \right). \quad (28)$$

Separating reals and imaginaries, we find

$$\left. \begin{aligned} x &= e^{\Phi/a} \cos \frac{\Psi}{a} \\ y &= e^{\Phi/a} \sin \frac{\Psi}{a} \end{aligned} \right\} \dots \dots \dots (29)$$

\* The equation  $\frac{d\Omega}{dz} = \frac{\partial \Omega}{\partial x}$  follows from  $\frac{\partial \Omega}{\partial x} = \frac{d\Omega}{dz} \frac{\partial z}{\partial x}$  because of  $\frac{\partial z}{\partial x} = 1$ .

By eliminating  $\Psi$  or  $\Phi$ , these equations give

$$x^2 + y^2 = e^{2\Psi/a} \quad \text{or} \quad \Phi = a \log r, \quad . . . (30)$$

and

$$\frac{y}{x} = \tan \frac{\Psi}{a}. \quad . . . . . (31)$$

The equipotential curves  $\Phi = \text{const.}$  are circles about  $O$ ; the stream lines are therefore straight lines radiating from  $O$  (fig. 1, *a*). This flow thus represents the two-dimensional analogue of the spherically symmetric flow from a source, considered on p. 199. The constant  $a$  is determined by the amount of fluid passing through unit length of a circular cylinder whose axis contains the source.

On the other hand, if we interpret the curves  $\Psi = \text{const.}$  as equipotentials, then the circles are stream lines (fig. 1, *b*). This flow pattern is the prototype of a *circulation*. If the point  $O$  is isolated from the

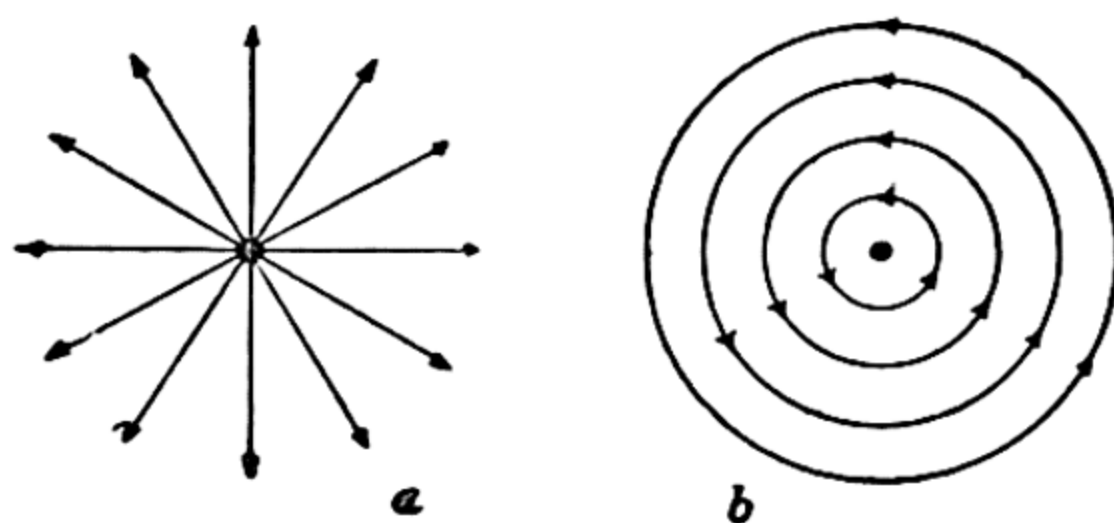


Fig. 1

field of flow by circumscribing a small circle, then  $\text{curl } \mathbf{v}$  is zero in all the remaining space. On the other hand, the line integral  $\oint \mathbf{v} d\mathbf{s}$  taken along one of the circles (or more generally, along a curve enclosing  $O$ ), no longer vanishes, for  $d\mathbf{s}$  always has the direction of  $\mathbf{v}$  along the circles, and all contributions to the integral are positive. We shall return to this type of flow in § 5 below.

We now consider the two-dimensional analogue of the problem of the sphere referred to above (p. 199), viz. the case of a circular cylinder immersed in a parallel stream. We take a co-ordinate system such that the undisturbed flow is parallel to the  $x$ -axis, and look for a function  $\Omega = f(z)$  which allows the entire  $\Omega$ -plane to be mapped on that portion of the  $z$ -plane which lies outside a circle of radius  $R$ . It is further required that the normal components of the velocity vanish at the surface of the cylinder—i.e. that the circle itself be a stream line, for which  $\Psi = \text{const.}$ —and finally, that the complex velocity  $w = d\Omega/dz$  shall pass over into the real quantity  $v = |\mathbf{v}|$  for large values of  $z$ . As will be shown immediately, the function

$$\Omega = v \left( z + \frac{R^2}{z} \right) \quad . . . . . (32)$$



fulfils these conditions. This is equivalent to

$$\Phi + i\Psi = vx \left(1 + \frac{R^2}{x^2 + y^2}\right) + ivy \left(1 - \frac{R^2}{x^2 + y^2}\right). \quad (32')$$

By solving the equation for  $z$ , it will be seen that for every value of  $\Omega$  there is one, and only one value of  $z$  whose modulus is greater than  $R$ ; hence there is a unique correspondence between the points of the  $\Omega$ -plane and the portion of the  $z$ -plane outside the circle. Moreover, we see that the curve  $\Psi = 0$  gives the circle  $x^2 + y^2 = R^2$  and the straight line  $y = 0$ , i.e. the circle coincides with a stream line. Finally, the value of  $\Omega$  becomes  $vz$  for  $z$  large, and so the complex velocity  $w = d\Omega/dz$  becomes equal to the real quantity  $v$ .

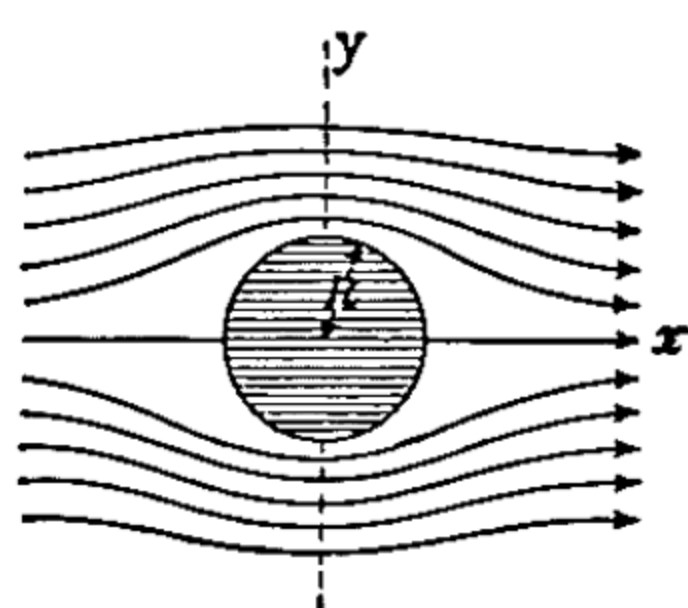


Fig. 2

Because of the complete symmetry of the flow pattern (fig. 2) in front of and behind the obstacle, we see that the magnitude of the complex velocity  $w$  is the same at any two points on the circle which are symmetric with respect to the  $y$ -axis. From Bernoulli's equation, it follows that the pressures are the same at two such points, so that the resultant of all the pressure forces vanishes. The concept of a simple potential flow thus leads to a paradoxical result in

this case. The reasons for this have been given above (p. 198).

#### 4. General Theorems concerning Vortex Motion and Circulation.

##### (a) *Distinction between vorticity and circulation*

We call the line integral  $\oint \mathbf{v} \cdot d\mathbf{s}$ , taken along a closed curve  $C$ , the circulation  $K$  along the curve. In view of Stokes's theorem,

$$\oint \mathbf{v} \cdot d\mathbf{s} = \int (\text{curl } \mathbf{v}) \cdot d\mathbf{S},$$

we might at first sight expect that in a flow where  $\text{curl } \mathbf{v}$  is zero everywhere the circulation along every curve would have to vanish. But, as has already been seen in the example of circular flow about a point  $O$ , where  $O$  itself is excluded (fig. 1, *b* above), the circulation may, in certain cases, differ from zero even when the flow is irrotational. This contradiction arises because the region is no longer simply connected after we cut out the point  $O$ , i.e. there now exist closed curves which cannot be closed in to a point without passing out of the region. These are all curves which enclose the small circle drawn about  $O$ . In the proof given of Stokes's theorem (p. 27), the simple connectivity of the

$\mathbf{v}$ -space was tacitly assumed. Hence, in the two-dimensional case, if we make the region multiply connected by excluding a portion of the plane, the circulation can be different from zero, despite the fact that the flow is irrotational in the region under consideration. Stokes's theorem may be retained, however, if the excluded region is replaced by suitably disposed vortices, i.e. if we take a certain distribution of  $\text{curl } \mathbf{v}$  on the boundary of the region. This procedure is especially important in the three-dimensional case.

(b) *Thomson's Theorem on the conservation of circulation*

Consider the particles of the fluid which form a closed curve  $C_0$  at time  $t = 0$ . Since we rule out discontinuities in the body of the fluid, particles originally adjacent will remain so as time goes on, i.e. these particles will continue to form a closed curve which, however, may undergo changes in shape. The location of a particle on a curve  $C$  may be given by a single parameter  $\lambda$ , say the length of arc. We now write the circulation along  $C_0$ :

$$K = \oint_{C_0} \mathbf{v} ds = \oint_{C_0} \mathbf{v} \frac{d\mathbf{r}}{d\lambda} d\lambda. \quad . . . . (33)$$

Differentiate with respect to the time. Since  $\lambda$  has nothing to do with time, the order of differentiation with respect to time and with respect to  $\lambda$  may be reversed, and we have\*

$$\begin{aligned} \frac{dK}{dt} &= \oint_{C_0} \frac{d\mathbf{v}}{dt} \frac{d\mathbf{r}}{d\lambda} d\lambda + \oint_{C_0} \mathbf{v} \frac{d}{d\lambda} \left( \frac{d\mathbf{r}}{dt} \right) d\lambda \\ &= \oint_{C_0} \frac{d\mathbf{v}}{dt} \frac{d\mathbf{r}}{d\lambda} d\lambda + \oint_{C_0} \frac{d}{d\lambda} \left( \frac{1}{2} \mathbf{v}^2 \right) d\lambda. \quad . . . (34) \end{aligned}$$

We now assume that the external forces are conservative, i.e.

$$\frac{d\mathbf{v}}{dt} = -\text{grad } U - \frac{1}{\rho} \text{grad } p = -\text{grad } [U + f(p)] = -\text{grad } V.$$

Then, since  $(\text{grad } V) \frac{d\mathbf{r}}{d\lambda} = \frac{dV}{d\lambda}$ ,

$$\frac{dK}{dt} = - \oint_{C_0} \frac{dV}{d\lambda} d\lambda + \oint_{C_0} \frac{d}{d\lambda} \left( \frac{1}{2} \mathbf{v}^2 \right) d\lambda.$$

Inasmuch as the integrand functions  $V$  and  $\frac{1}{2} \mathbf{v}^2$  are definite functions

\* Although two variables  $t$  and  $\lambda$  occur here, total derivatives are written, since the partial derivatives  $\partial/\partial t$  were used in another sense in § 2 (p. 195) above. Here we are concerned with the total change of  $\mathbf{v}$  in time, not merely the change occurring when the position is held fixed.

of position, the integrals taken along a closed curve  $C_0$  vanish, and we have

$$\frac{dK}{dt} = 0 \quad \text{or} \quad K = \text{const.} \quad . \quad . \quad . \quad (35)$$

*The circulation is constant for a closed curve composed of a given set of fluid particles if the external forces are conservative. This result was given by Sir W. Thomson (Lord Kelvin).*

### (c) Helmholtz's Vortex Theorems

A number of important theorems, found by Helmholtz in another way, may be derived from the Thomson theorem. Imagine the field lines of the vector  $\text{curl } \mathbf{v}$  drawn, i.e. the curves whose tangents at any point of space give the direction of  $\text{curl } \mathbf{v}$ , or the axis of rotation  $\omega$  of the fluid particles. In order to obtain a measure of the strength as well, we assume that a number of field lines proportional to  $\text{curl } \mathbf{v}$  pass through every square centimetre of a surface normal to  $\text{curl } \mathbf{v}$ . These field lines will now be called *vortex lines*.

From the general theorem of vector analysis

$$\text{div curl } \mathbf{v} = 0,$$

we conclude that *the vortex lines can neither begin nor terminate within the fluid; thus they are either closed curves or, in the case of a bounded fluid, they begin and end on the surface.*

A tube-like surface made up of vortex lines is called a *vortex tube*. If the cross-section of this structure is so small that  $\text{curl } \mathbf{v}$  may be considered constant over it, we speak of a *vortex filament*. For such vortex filaments we have the theorem that the product of the cross-section by the magnitude of  $\text{curl } \mathbf{v}$ —the so-called *vortex strength*—is constant along the filament; for if we take the surface integral in (a) over a portion of a vortex filament bounded by two cross-sections  $A_1$  and  $A_2$  normal to  $\text{curl } \mathbf{v}$ , the contribution of the lateral surface drops out, since  $d\mathbf{S}$  is normal to  $\text{curl } \mathbf{v}$  everywhere on this surface, and there remains only the contribution of the two ends  $A_1$  and  $A_2$ . Since no vortex lines can originate on the inside, and since  $\text{curl } \mathbf{v}$  has the same direction as  $d\mathbf{S}$  at one end, and the opposite direction at the other end,

$$|\text{curl } \mathbf{v}|_1 A_1 = |\text{curl } \mathbf{v}|_2 A_2. \quad . \quad . \quad . \quad (36)$$

Further, we can show that every vortex filament always consists of the same fluid particles, i.e. the filament travels along with the fluid. Draw a closed curve  $C$  (fig. 3) which does not encircle the vortex



tube. By Stokes's theorem, the circulation  $\oint_C \mathbf{v} \cdot d\mathbf{s}$  along this curve is

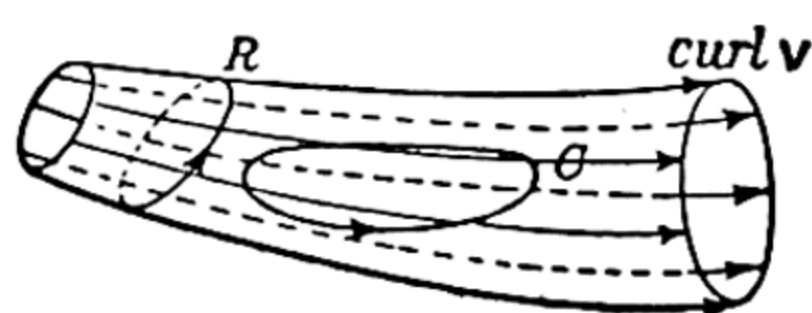


Fig. 3

zero, since  $\text{curl } \mathbf{v}$  remains normal to  $d\mathbf{s}$  at all points of the surface enclosed by  $C$ . According to Thomson's theorem, the circulation along a curve composed of the particles which constitute the curve  $C$  at time  $t$  must remain zero. This means that the curve  $C$  remains on the

surface of the vortex tube. Since these considerations are valid for every curve on the lateral surface which does not encircle the tube, we may imagine the entire surface made up of elements bounded by such curves, the circulation being zero along their boundaries. This means, however, that *the vortex tube consists always of the same fluid particles*.

This law holds for vortex filaments also, since a filament may be thought of as the common element of two vortex tubes. The motion of vortex filaments can be observed, for example, with smoke rings produced by blowing tobacco smoke from the mouth. The smoke particles serve to make the motion of the air visible. The vortex ring moves onward with the air expelled from the mouth. Rotation of the air around the axis of the ring may be observed only in the first stages of the motion, since it is soon destroyed by the viscosity of the air, which has been left out of account here.

We now consider a curve  $R$  which encircles the vortex filament. The circulation along this curve is equal to the vortex strength of the filament, by Stokes's theorem. Now by Thomson's theorem, the circulation round this fluid curve is constant in time, so that the vortex strength must be constant also. We thus obtain the theorem: *In a frictionless fluid, the strengths of the vortex filaments are constant in time, and hence vortices can be neither created nor destroyed.*

The remarkable properties of circular vortex rings—their indestructibility, indivisibility, &c.—led Sir W. Thomson to formulate the ingenious theory that atoms are vortices in the ether. This idea was developed by Sir J. J. Thomson in an Adams Prize Essay (1883). The matter is only of historical interest now.

(d) *The law of Biot and Savart for the flow generated by a vortex filament*

If the vector  $\text{curl } \mathbf{v}$  is given as a function of position, the field of flow may be calculated from it. We take the simple case of a single vortex filament of strength  $K$ . This problem corresponds exactly to the calculation of the electromagnetic field of a thin wire carrying a current  $I$ , since for a curve of area  $A$  encircling the vortex filament,











The second integral on the right vanishes, however, for a plane curve, as may be shown readily. Thus, if we call the unit tangent vector  $t$ , and the unit vector normal to the plane of the curve  $b$ , then  $n = [bt]$ , whence

$$\oint kn ds = k \oint [(bt)] ds = k [b \oint t ds] = k [b \oint ds] = 0, \quad (45)$$

since  $\oint ds$  vanishes for every closed curve. In order to evaluate the first integral, it is convenient to introduce a complex number

$$Z = F_y + iF_x, \quad . \quad . \quad . \quad . \quad . \quad (46)$$

whose radius vector in the complex plane is thus a reflection of the force vector in the line bisecting the first quadrant. Now, by equations (43) and (44),

$$F_y = -\frac{\rho}{2} \oint v^2 \cos(ny) ds, \quad F_x = -\frac{\rho}{2} \oint v^2 \cos(nx) ds. \quad (47)$$

But from fig. 4 we see that

$$ds \cos(ny) = dx, \quad ds \cos(nx) = -dy. \quad . \quad . \quad (48)$$

We thus obtain

$$\begin{aligned} Z &= -\frac{\rho}{2} \oint v^2 (dx - i dy) \\ &= -\frac{\rho}{2} \oint (v_x^2 + v_y^2) (dx - i dy). \quad . \quad . \quad (49) \end{aligned}$$

If we add under the integral sign the expression

$$2i(v_x dy - v_y dx) (v_x - i v_y),$$

which is zero, by equation (39') (p. 208), we obtain

$$Z = -\frac{\rho}{2} \oint (v_x^2 - 2i v_x v_y - v_y^2) (dx + i dy) = -\frac{\rho}{2} \oint w^2 dz. \quad (50)$$

Since the disturbance caused by the obstacle vanishes at a great distance from it, the complex velocity  $w$  may be represented, according to equation (41), by a series of descending powers of  $z$ , if we take the origin within the obstacle:

$$w = w_\infty + \frac{a_{-1}}{z} + \frac{a_{-2}}{z^2} + \dots \quad . \quad . \quad . \quad (41)$$

The corresponding complex potential is, by integration,

$$\Omega = a_0 + w_\infty z + a_{-1} \log z - \frac{a_{-2}}{z} + \dots \quad . \quad . \quad (41')$$

By equation (42), the coefficient  $a_{-1}$  is connected with the circulation by the equation

$$a_{-1} = \frac{K}{2\pi i}.$$

If we form the integral  $\oint w^2 dz$ , we are concerned only with the term in  $1/z$ . In squaring  $w$ , this term obtains the coefficient  $2w_\infty a_{-1}$  and the value of the integral becomes

$$\oint w^2 dz = 2\pi i \frac{2K}{2\pi i} w_\infty, \quad . . . . . (51)$$

whence, by (50),

$$Z = -\rho K w_\infty = -\rho K (v_{x_\infty} - i v_{y_\infty}). \quad . . . (52)$$

From the definition of  $Z$ , in (46), we thus have

$$\left. \begin{aligned} F_x = I(z) &= +\rho K v_{y_\infty} \\ F_y = R(z) &= -\rho K v_{x_\infty} \end{aligned} \right\} . . . . . (53)$$

These equations constitute the Kutta-Joukowski Lift Formula, so named after its discoverers. It shows at once that the simple potential flow about a circular cylinder causes no resultant force on the cylinder, since the complex velocity

$$w = v - \frac{vR^2}{z^2},$$

which, by equation (32) (p. 202) occurs in that case, contributes no term in  $1/z$  to the integral  $\oint w^2 dz$ . In reality, however, the breaking away of the boundary layer from the cylinder causes a vortex region to form behind the body, and this invalidates our calculations. The turbulent detachment of the boundary layer can be checked by setting the cylinder into rotation. In this case we have, superimposed upon the solution (32) (p. 202), another flow which streams around the cylinder in circular paths. The corresponding complex potential has already been encountered in § 3 (p. 202), and is  $A \log z$ . The complete flow is then given by the potential function

$$\Omega = v \left( z + \frac{R^2}{z} \right) + A \log z, \quad . . . . . (54)$$

and the complex velocity

$$w = v - \frac{vR^2}{z^2} + \frac{A}{z}. \quad . . . . . (55)$$

The circulation is

$$K = 2\pi i A, \quad \text{i.e. } A = \frac{K}{2\pi i}. \quad . . . . . (56)$$

If we assume complete adhesion of the boundary layer, the circulation has the value  $2\pi\omega R^2$ , where  $\omega$  is the angular velocity of rotation of the cylinder and  $R$  its radius. If the original flow is along the  $x$ -axis, we have a force in the direction of the  $y$ -axis,

$$F_y = -2\pi\rho\omega R^2v. \quad . \quad . \quad . \quad (57)$$

The force is directed toward the side on which the stream and peripheral velocities have the same direction, since  $K$  is positive when the surface is on the left.\* This phenomenon is known as the Magnus Effect, after its discoverer. This somewhat unexpected transverse force has recently had an interesting practical application in the rotor ship.

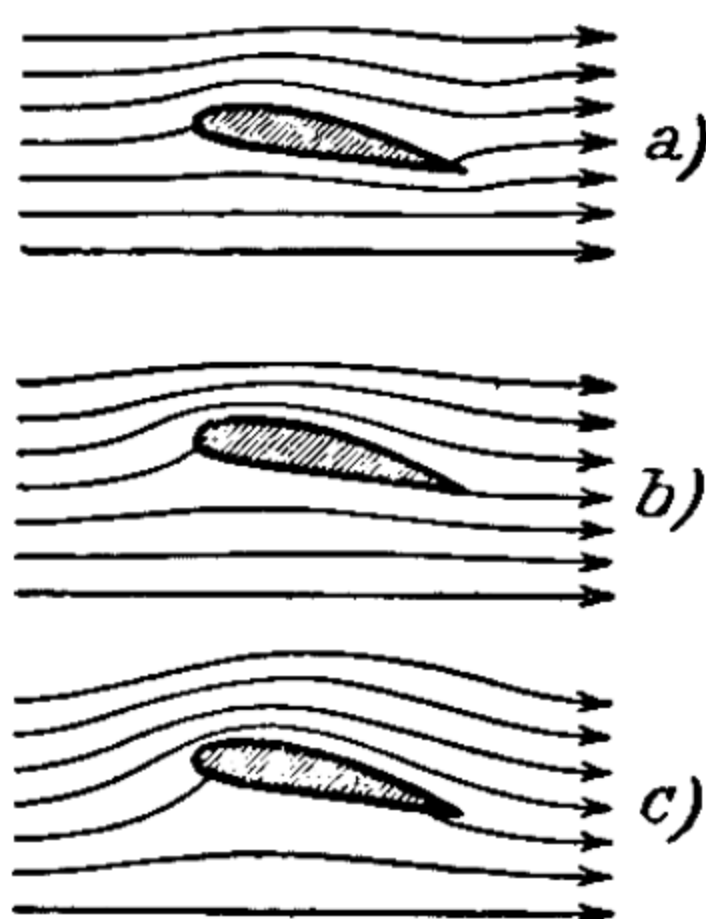


Fig. 5

By means of relatively simple transformation functions the flow pattern of a combined parallel flow and circulation about a circular cylinder may be converted into the flow about a profile similar to the wing sections used in aeronautics. According to the magnitude of the circulation about the cylinder, we obtain flow patterns of types *a* or *c* (fig. 5). Case *b* represents the transition. In case *a*, where the circulation is small, the flow would have to pass around the trailing edge to the upper surface; in case *c*, the flow would be required to leave from the lower surface. Since it is an empirical fact that a fluid does not pass around sharp corners in this way, the circulation which actually occurs must be that corresponding to case *b*, where the fluid leaves the profile smoothly at the trailing edge. Thus the lift of a given wing section may be calculated in advance if the section is obtainable from the circle by conformal transformation. It is to be remembered, however, that the two-dimensional computation is a somewhat rough approximation to the actual case.

It may be of interest to see how the creation of circulation about the profile is to be reconciled with Thomson's theorem concerning the constancy of the circulation. If we imagine the wing section brought into an ordinary parallel stream for which no circulation exists for any curve, then no circulation can come about, according to Thomson's theorem. But, as already emphasized, the friction at the boundary layer cannot be neglected, even for a fluid of low viscosity. As a consequence, Thomson's theorem may be applied only to a fluid curve enclosing the body, and at a great distance from it. The circulation remains zero for such a curve. To the same degree as a circulation develops about the profile, an oppositely directed vortex—the so-called *starting vortex*†—develops

\* The value in (57) holds only in the limiting case where no detachment of the boundary layer occurs. Actually, however, the value of the rotational velocity of the fluid to be substituted in the equation is only about  $\omega/2$ , since a partial detachment of the boundary layer always takes place.

† Ger. *Anfahrwirbel*.



at the trailing edge. When its strength becomes equal to the final value of the circulation about the profile, this vortex is shed downstream, and the flow pattern remains steady thereafter. On account of the presence of both this vortex and the circulation about the profile, the circulation remains zero about a curve which encloses the profile and is at a great distance from it, since it encloses both the vortex and the circulation about the profile.

## 6. Undulatory Propagation of Disturbances in Fluids. (Sound Waves.)

Compressions and expansions which are propagated in liquids and gases in the form of waves are among the non-stationary kinetic phenomena lending themselves most readily to mathematical treatment. The reason is that the velocities and density changes which occur are so small that products of these quantities may be neglected. For this case, then, we may strike out the term  $\mathbf{v} \text{ grad. } \mathbf{v}$  from the equations of motion (p. 195). In the absence of external forces these equations then become

$$\frac{\partial \mathbf{v}}{\partial t} = -\frac{1}{\rho} \text{grad } p. \quad . \quad . \quad . \quad . \quad . \quad (58)$$

Let the density be represented by

$$\rho = \bar{\rho}[1 + \sigma(t, x, y, z)], \quad . \quad . \quad . \quad . \quad . \quad (59)$$

where  $\bar{\rho}$  is the average density and  $\sigma$  is a measure of the departures from this normal value. Putting this expression into the equation of continuity (12) (p. 196), and neglecting terms of the second order, we obtain after division by  $\bar{\rho}$ :

$$\frac{\partial \sigma}{\partial t} + \text{div } \mathbf{v} = 0. \quad . \quad . \quad . \quad . \quad . \quad (60)$$

Since (58) is equivalent to three scalar equations, we have four equations for the five variables  $v_x, v_y, v_z, p, \sigma$ ; thus we need one more equation. This comes from the relation between  $p$  and  $\rho$ , or between  $p$  and  $\sigma$ . For gases, Boyle's Law

$$\frac{p}{\bar{p}} = \frac{\rho}{\bar{\rho}}$$

( $\bar{p}$  = average pressure) suggests itself. However, this leads to erroneous values for the velocity of sound, the reason being that Boyle's Law holds for constant temperature, while the compressions and expansions follow one another so rapidly here that the heat generated according to the laws of thermodynamics (cf. p. 506) cannot be dissipated sufficiently rapidly. We must apply the adiabatic law (p. 506), which holds for complete thermal insulation. In terms of the density, which is inversely proportional to the volume, this law is

$$\frac{p}{\bar{p}} = \left(\frac{\rho}{\bar{\rho}}\right)^{\gamma} = (1 + \sigma)^{\gamma} = 1 + \gamma\sigma, \quad . . . (61)$$

where  $\gamma$  is the ratio of the specific heat at constant pressure to that at constant volume. Then

$$\text{grad } p = \bar{p} \gamma \text{ grad } \sigma, \quad . . . (62)$$

and, again neglecting products of small quantities,

$$\frac{1}{\rho} \text{grad } p = \frac{\bar{p} \gamma}{\bar{\rho}} \text{grad } \sigma. \quad . . . (63)$$

If we substitute this value in the simplified hydrodynamical equation (58), we obtain

$$\frac{\partial v}{\partial t} = - \frac{\bar{p} \gamma}{\bar{\rho}} \text{grad } \sigma. \quad . . . (64)$$

It is easy to eliminate  $v$  from equations (60) and (64). Since derivatives with respect to position and time may be taken in either order, we need only differentiate equation (60) once more with respect to  $t$ , take the divergence of equation (64), and equate the two resulting expressions for

$$\frac{\partial}{\partial t} (\text{div } v).$$

The result is 
$$\text{div grad } \sigma = \Delta \sigma = \frac{\rho}{\bar{p} \gamma} \frac{\partial^2 \sigma}{\partial t^2}. \quad . . . (65)$$

But according to p. 63, this is the differential equation of a wave motion propagated with the wave velocity

$$v = \sqrt{\frac{\bar{p} \gamma}{\bar{\rho}}}. \quad . . . (66)$$

Thus we have the formula for the velocity of sound in gases. The average density  $\bar{\rho}$  is still a function of the temperature. By p. 494, we have

$$\bar{\rho} = \frac{\rho_0}{1 + \alpha \theta}, \quad . . . (67)$$

where  $\rho_0$  is the density at  $0^\circ \text{C.}$ ,  $\theta$  is the centigrade temperature, and  $\alpha$  is the coefficient of expansion of the gas. Hence

$$v = \sqrt{\frac{\bar{p} \gamma (1 + \alpha \theta)}{\rho_0}}. \quad . . . (68)$$

Since  $\rho_0$  is proportional to  $\bar{p}$ ,  $v$  is independent of the pressure.

*Ex. 62.* Calculate the velocity of sound in water, the value of the compressibility being  $\kappa = 50 \times 10^{-6} \text{ atm.}^{-1}$  and  $\gamma \doteq 1$ .

## 7. Hydrodynamics of Viscous Fluids.

### (a) Simple linear laminar flow; Poiseuille's Law

The term "laminar" is applied to a flow pattern in which the fluid may be divided into parallel layers which flow past one another with various velocities. The simplest case is that where the velocities of all layers have the same direction, say that of the  $x$ -axis. As a result of internal friction (viscosity), the more rapidly moving layer tends to drag the neighbouring layer along with it, and thus accelerates it. Reciprocally, the slower layer tends to retard the faster one. The correct definition of this force goes back to Newton. The force is proportional to the common area of contact  $S$  of the two layers and to the velocity gradient normal to the direction of flow. Thus if the velocity  $v$  increases along the positive  $z$ -axis, the layer just above that for which  $z = 0$  exerts a tangential force upon an area  $S$  of the latter of amount

$$F = \eta S \frac{dv}{dz}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (69)$$

In this case the force is in the direction of the positive  $x$ -axis, and thus is to be reckoned positive. The factor of proportionality  $\eta$  is called the *coefficient of viscosity*. Laminar flow occurs, for example, when a viscous fluid flows slowly through a circular cylindrical tube. From experience we know that the liquid adheres to the walls; hence the velocity increases as we go toward the axis of the tube, and  $dv/dr$  is negative. Take the element of volume to be a hollow cylinder of fluid with radii  $r$  and  $r + dr$ . The adjacent layer on the inside exerts a force on this cylinder which is along the positive  $x$ -axis and amounts to

$$F_1 = -2\pi\eta l r \frac{dv}{dr},$$

where  $l$  is the length of the tube. On account of the decrease of velocity toward the outside, the contiguous layer of fluid on the outside tends to retard the cylinder with a force

$$F_2 = +2\pi\eta l (r + dr) \left( \frac{dv}{dr} + \frac{d^2v}{dr^2} dr \right).$$

If the pressure at the inlet end of the tube is  $p_1$  and that at the outlet  $p_0$ , then the resultant hydrostatic force on the ends of the volume element is  $2\pi r dr (p_1 - p_0)$ . In the stationary state corresponding to uniform motion the pressure force and the forces of viscosity must be in equilibrium. Hence, neglecting higher order terms,

$$2\pi\eta l \frac{dv}{dr} + 2\pi\eta l r \frac{d^2v}{dr^2} = 2\pi\eta l \frac{d}{dr} \left( r \frac{dv}{dr} \right) = -2\pi r (p_1 - p_0). \quad (70)$$



This equation may be integrated at once:

$$\frac{dv}{dr} = -\frac{(p_1 - p_0)r}{2\eta l} + \frac{C}{r}. \quad \cdot \cdot \cdot \cdot (71)$$

Integrating a second time,

$$v = -\frac{(p_1 - p_0)}{2\eta l} \frac{r^2}{2} + C \log r + D. \quad \cdot \cdot \cdot (72)$$

The constants of integration  $C$  and  $D$  are determined in the following way: Since  $v$  must remain finite along the axis ( $r = 0$ ),  $C$  must be zero. Moreover,  $v$  must vanish at the wall, where  $r = a$ . Hence,

$$D = \frac{p_1 - p_0}{4\eta l} a^2.$$

The solution is thus 
$$v = \frac{p_1 - p_0}{4\eta l} (a^2 - r^2). \quad \cdot \cdot \cdot \cdot (73)$$

If the velocity vectors of individual fluid elements are drawn with their initial points in a given cross-section, their termini lie on a surface of rotation whose meridian section is a parabola. In this case we speak of a *parabolic velocity profile*. The total volume  $W$  issuing per second is obtained by taking the integral  $\int v dS$  over a cross-section. In this case we have

$$W = \int_0^a \frac{p_1 - p_0}{4\eta l} (a^2 - r^2) 2\pi r dr = \frac{\pi (p_1 - p_0) a^4}{8\eta l}. \quad (74)$$

This is Poiseuille's Formula, which states that the quantity of fluid issuing each second is directly proportional to the pressure difference and to the fourth power of the radius of the tube, and inversely proportional to the length of the tube and to the coefficient of viscosity.

Simple laminar flow, however, is not the only form which can develop in a tube. At greater velocities the flow breaks down into another state where (1) the velocity of efflux is no longer proportional to the pressure difference, but increases more slowly, (2) the values of velocity and pressure at a given point are not constant in time, as before, but oscillate about mean values, and (3) the paths of individual particles are no longer rectilinear; there is a continual intermixing of fluid elements, as may easily be demonstrated by injecting filaments of coloured liquid. This state of flow is called *turbulence*. The initiation of turbulence is determined by the value of a dimensionless quantity called *Reynolds's number*, after its discoverer. This quantity, for a circular tube of radius  $a$ , is given by

$$R = \frac{2\rho a v}{\eta}, \quad \cdot \cdot \cdot \cdot (75)$$

where  $\rho$  is the density of the fluid and  $v$  is its velocity. There is a rather definite lower limit of the value of  $R$  corresponding to the beginning of turbulence. Turbulent flow has not been produced for values of  $R$  below about 2200. Whether or not turbulence appears for values of  $R$  above 2200 depends upon the degree to which the fluid is free of disturbances, especially at the inlet. The theory of turbulence and of the breakdown of laminar flow is still an incompletely solved problem of hydrodynamics.

(b) *Extension to arbitrary states of flow*

In order to calculate the stresses on a volume element of a viscous fluid we can avail ourselves of the formulæ of the theory of elasticity, provided we bear in mind that in place of the displacements  $u, v, w$  (p. 162), their time derivatives appear, since the viscous forces operate only when motion occurs. But these derivatives are merely the components of the flow velocity. Thus we must put into equations (21) (p. 166),

$$\left. \begin{aligned} e_{11} &= \frac{\partial v_x}{\partial x} & e_{12} &= \frac{\partial v_x}{\partial y} + \frac{\partial v_y}{\partial x} = e_{21} \\ e_{22} &= \frac{\partial v_y}{\partial y} & e_{23} &= \frac{\partial v_y}{\partial z} + \frac{\partial v_z}{\partial y} = e_{32} \\ e_{33} &= \frac{\partial v_z}{\partial z} & e_{31} &= \frac{\partial v_z}{\partial x} + \frac{\partial v_x}{\partial z} = e_{13} \end{aligned} \right\} \cdot \cdot \cdot \quad (76)$$

We shall assume, in what follows, that the fluid is incompressible; hence  $e_{11} + e_{22} + e_{33} = 0$ . We must now introduce the coefficient of viscosity  $\eta$  in place of the elastic constants  $E$  and  $\sigma$ . To see how this is to be done, we return to the above special case of linear laminar flow. Let the flow now be along the  $y$ -axis, and let the velocity increase in the direction of the positive  $x$ -axis. Thus

$$v_x = 0, \quad v_y = v, \quad v_z = 0. \quad \cdot \cdot \cdot \quad (77)$$

The elastic equations of p. 174 give the shearing stress on unit area normal to the  $x$ -axis, acting in the direction of the  $y$ -axis, as

$$P_{12} = \frac{E}{2(1 + \sigma)} e_{12} = \frac{E}{2(1 + \sigma)} \frac{dv}{dx}, \quad \cdot \cdot \cdot \quad (78)$$

while equation (69) (p. 214) gives

$$\eta \frac{dv}{dx}$$

for this quantity. Thus  $\frac{E}{2(1+\sigma)}$

is to be replaced by  $\eta$ , and we have

$$\left. \begin{aligned} P_{11} &= 2\eta \frac{\partial v_x}{\partial x} & P_{12} &= \eta \left( \frac{\partial v_x}{\partial y} + \frac{\partial v_y}{\partial x} \right) = P_{21} \\ P_{22} &= 2\eta \frac{\partial v_y}{\partial y} & P_{23} &= \eta \left( \frac{\partial v_y}{\partial z} + \frac{\partial v_z}{\partial y} \right) = P_{32} \\ P_{33} &= 2\eta \frac{\partial v_z}{\partial z} & P_{31} &= \eta \left( \frac{\partial v_z}{\partial x} + \frac{\partial v_x}{\partial z} \right) = P_{13} \end{aligned} \right\} \quad (79)$$

According to p. 170, the components of the resultant of the stress forces per cubic centimetre become

$$\begin{aligned} F_x &= \frac{\partial P_{11}}{\partial x} + \frac{\partial P_{21}}{\partial y} + \frac{\partial P_{31}}{\partial z} = \text{div } \mathbf{P}_1 = 2\eta \frac{\partial e_{11}}{\partial x} + \eta \frac{\partial e_{21}}{\partial y} + \eta \frac{\partial e_{31}}{\partial z} \\ &= \eta \left( 2 \frac{\partial^2 v_x}{\partial x^2} + \frac{\partial^2 v_x}{\partial y^2} + \frac{\partial^2 v_y}{\partial x \partial y} + \frac{\partial^2 v_z}{\partial x \partial z} + \frac{\partial^2 v_x}{\partial z^2} \right), \quad \dots \quad (80) \end{aligned}$$

$$\text{or} \quad F_x = \eta (\Delta v_x + \frac{\partial}{\partial x} \text{div } \mathbf{v}) = \eta \Delta v_x, \quad \dots \quad (81)$$

since  $\text{div } \mathbf{v} = 0$ . Similarly,

$$\left. \begin{aligned} F_y &= \eta \Delta v_y \\ F_z &= \eta \Delta v_z \end{aligned} \right\} \quad \dots \quad (81')$$

If these forces are added to the previously considered forces acting on unit volume of a fluid, we obtain the additional term  $\eta \Delta \mathbf{v}$  in the vector formulation. But by the familiar vector formula

$$\text{curl curl } \mathbf{v} = \text{grad div } \mathbf{v} - \Delta \mathbf{v},$$

the viscosity term  $\eta \Delta \mathbf{v}$  becomes simply  $-\eta \text{curl curl } \mathbf{v}$ , since  $\text{div } \mathbf{v} = 0$ . Denoting, as above, the external force per gramme by  $\mathbf{G}$ , the fundamental hydrodynamical equation for incompressible viscous fluids becomes

$$\frac{\partial \mathbf{v}}{\partial t} + \mathbf{v} \text{grad} \cdot \mathbf{v} = \mathbf{G} - \frac{1}{\rho} \text{grad } p - \frac{\eta}{\rho} \text{curl curl } \mathbf{v}. \quad (82)$$

If the stresses due to viscosity are allowed to remain in the equation, this expression becomes, using (80),

$$\frac{\partial \mathbf{v}}{\partial t} + \mathbf{v} \text{grad} \cdot \mathbf{v} = \mathbf{G} - \frac{1}{\rho} \text{grad } p + \frac{1}{\rho} (i \text{div } \mathbf{P}_1 + j \text{div } \mathbf{P}_2 + k \text{div } \mathbf{P}_3). \quad (83)$$



(c) *Stokes's Law for the fluid resistance to a moving sphere*

Stokes was the first to succeed in determining the resistance experienced by a sphere moving uniformly through a viscous fluid. Naturally, it amounts to the same thing if we consider the sphere fixed and the fluid to be flowing by it with the opposite velocity  $\mathbf{v}_0$ . We are thus concerned with finding a special solution of (82) which corresponds to the flow about a spherical obstacle and calculating the resultant force acting on the sphere from the flow pattern obtained. Stokes assumed the speed to be so small that the expression  $\mathbf{v} \text{ grad} \cdot \mathbf{v}$ , containing the products

$$v_x \frac{\partial v_x}{\partial x}, \dots$$

could be neglected. That is, in the stationary state—which alone interests us—the left member of (82) may be set equal to zero. This approximation, however, leads to erroneous results in the case of a cylinder, as Stokes himself pointed out. As Oseen has shown, the term  $\mathbf{v} \text{ grad} \cdot \mathbf{v}$  must be taken into account, even for low speeds, at least to the extent that it is included in the form  $\mathbf{v}_0 \text{ grad} \cdot \mathbf{v}$ —i.e. the flow  $\mathbf{v}$  is replaced by the undisturbed flow  $\mathbf{v}_0$  except in the derivatives of velocity. For the sphere, this furnishes only a small correction term, Stokes's analysis being thus justified, and his result verified. For the sake of simplicity, we shall therefore give a brief reproduction of Stokes's mathematics. Since we are concerned only with viscous forces, we leave the external force—the force of gravity—out of consideration. Its only effect is to produce a hydrostatic buoyancy. We proceed, then, by replacing all forces not connected with viscosity by the hydrostatic forces, since the velocity is low. However, in computing the *total* force, these contributions must not be forgotten. Thus, in the problem of a sphere falling through a viscous liquid we must not forget the buoyant force. In order to calculate the viscous forces, we thus consider the fluid freed from gravity, and the hydrodynamic equations reduce to

$$\text{grad } p = \eta \Delta \mathbf{v}; \quad \text{div } \mathbf{v} = 0. \quad . \quad . \quad . \quad . \quad (84)$$

We try a solution for  $\mathbf{v}$  which is the sum of a potential flow  $\mathbf{v}_1 = \text{grad } \Phi$  and a flow  $\mathbf{v}_2$  not derivable from a potential. Then

$$\text{grad } p = \eta \Delta \text{grad } \Phi + \eta \Delta \mathbf{v}_2 = \eta \text{grad } \Delta \Phi + \eta \Delta \mathbf{v}_2. \quad (85)$$

We further assume that  $\mathbf{v}_2$  satisfies the equation  $\Delta \mathbf{v}_2 = 0$ . Integration of (85) then gives

$$p = \eta \Delta \Phi + p_0. \quad . \quad . \quad . \quad . \quad . \quad (86)$$

The condition of incompressibility,  $\text{div } \mathbf{v} = 0$ , furnishes the additional equation

$$\text{div grad } \Phi + \text{div } \mathbf{v}_2 = \Delta \Phi + \text{div } \mathbf{v}_2 = 0. \quad . \quad . \quad (87)$$

The simplest assumption we can make regarding  $\mathbf{v}_2$  is that there is only one component  $v_{2x}$  different from zero. This quantity must then satisfy the equation  $\Delta v_{2x} = 0$ . The simplest solution of this equation, corresponding to spherical symmetry, is

$$v_{2x} = \frac{a}{r}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (88)$$

Thus we put

$$\mathbf{v}_2 = \frac{a}{r} \mathbf{i}, \quad \text{div } \mathbf{v}_2 = \mathbf{i} \left( \text{grad } \frac{a}{r} \right) = a \frac{\partial}{\partial x} \frac{1}{r}. \quad . \quad . \quad (89)$$

According to (87), this gives the following differential equation for  $\Phi$ :

$$\Delta \Phi + a \frac{\partial}{\partial x} \frac{1}{r} = 0. \quad . \quad . \quad . \quad . \quad . \quad . \quad (90)$$

We may verify in a moment that a particular integral of this equation is

$$\Phi = -\frac{a}{2} \frac{\partial r}{\partial x}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (91)$$

for this gives

$$\begin{aligned} \Delta \Phi &= -\frac{a}{2} \frac{\partial}{\partial x} \Delta r = -\frac{a}{2} \frac{\partial}{\partial x} \text{div grad } r = -\frac{a}{2} \frac{\partial}{\partial x} \text{div } \frac{\mathbf{r}}{r} \\ &= -\frac{a}{2} \frac{\partial}{\partial x} \left( \frac{3}{r} - \frac{r^2}{r^3} \right) = -a \frac{\partial}{\partial x} \frac{1}{r}. \end{aligned}$$

In order to have sufficient constants to satisfy the boundary conditions, we add to this solution functions which satisfy the equation  $\Delta \Phi = 0$ . Thus these functions drop out when substituted in equation (90). Since our previous particular solutions vanish at infinity, we must add a term which fulfils the condition that the flow at a great distance from the sphere is parallel to the  $x$ -axis. A suitable term is  $v_0 x$ . Another function which satisfies  $\Delta \Phi = 0$  is, for example,  $b/r$ . But further, on account of the interchangeability of the operators  $\partial/\partial x$  and  $\Delta$ , all derivatives of  $1/r$  also satisfy the equation  $\Delta \Phi = 0$ . We shall show presently that all the boundary conditions can be satisfied by adding the single function

$$b \frac{\partial}{\partial x} \frac{1}{r};$$

our assumed solution is thus

$$\mathbf{v} = \text{grad} \left( v_0 x + b \frac{\partial \frac{1}{r}}{\partial x} - \frac{a}{2} \frac{\partial r}{\partial x} \right) + \frac{a}{r} \mathbf{i}. \quad (92)$$

Written in component form this is

$$v_x = v_0 + b \frac{\partial^2 \frac{1}{r}}{\partial x^2} - \frac{a}{2} \frac{\partial^2 r}{\partial x^2} + \frac{a}{r} = v_0 + b \left( \frac{3x^2}{r^5} - \frac{1}{r^3} \right) - \frac{a}{2} \left( \frac{1}{r} - \frac{x^2}{r^3} \right) + \frac{a}{r}.$$

$$v_y = b \frac{\partial^2 \frac{1}{r}}{\partial x \partial y} - \frac{a}{2} \frac{\partial^2 r}{\partial x \partial y} = b \frac{3xy}{r^5} + \frac{a}{2} \frac{xy}{r^3}.$$

$$v_z = b \frac{\partial^2 \frac{1}{r}}{\partial x \partial z} - \frac{a}{2} \frac{\partial^2 r}{\partial x \partial z} = b \frac{3xz}{r^5} + \frac{a}{2} \frac{xz}{r^3}.$$

For  $r = R$  all components of  $\mathbf{v}$  must vanish for all values of  $x, y, z$ . This is true for the  $y$ - and the  $z$ -components if we put

$$b = -\frac{aR^2}{6}. \quad (93)$$

Then the coefficient of  $x^2$  in the  $x$ -component vanishes of itself, and by putting the remaining terms equal to zero we obtain

$$a = -v_0 \frac{3R}{2}, \quad \text{and} \quad b = \frac{v_0 R^3}{4}. \quad (94)$$

The solution corresponding to the boundary conditions is then

$$\left. \begin{aligned} v_x &= v_0 \left( 1 - \frac{3}{4} \frac{R}{r} - \frac{1}{4} \frac{R^3}{r^3} \right) - \frac{3}{4} \frac{v_0 R}{r^3} \left( 1 - \frac{R^2}{r^2} \right) x^2 \\ v_y &= -\frac{3v_0 R}{4r^3} \left( 1 - \frac{R^2}{r^2} \right) xy \\ v_z &= -\frac{3v_0 R}{4r^3} \left( 1 - \frac{R^2}{r^2} \right) xz \end{aligned} \right\}. \quad (95)$$

By equations (86), (90), and (94) the pressure becomes

$$p = p_0 + \eta \Delta \Phi = p_0 - a\eta \frac{\partial \frac{1}{r}}{\partial x} = p_0 - \frac{3v_0}{2} \frac{\eta R}{r^3} x. \quad (96)$$

The resultant force acting on the sphere is obtained by evaluating the stresses from the velocity components by equation (79) (p. 217), and



taking the integral of the stresses and of the pressures over the sphere. We thus have (cf. p. 168):

$$\mathbf{F} = - \oint p d\mathbf{S} + \oint \{ \mathbf{P}_1 \cos(\mathbf{n}i) + \mathbf{P}_2 \cos(\mathbf{n}j) + \mathbf{P}_3 \cos(\mathbf{n}k) \} d\mathbf{S}, \quad (97)$$

the pressure having the negative sign (see footnote, p. 191). On account of symmetry, the resultant can have only an  $x$ -component. The value of this is

$$F_x = F = - \oint p i dS + \oint \{ P_{11} \cos(\mathbf{n}i) + P_{21} \cos(\mathbf{n}j) + P_{31} \cos(\mathbf{n}k) \} dS. \quad (98)$$

Because of the symmetry of the stress tensor ( $P_{12} = P_{21}$ ), we have

$$\{ P_{11} \cos(\mathbf{n}i) + P_{21} \cos(\mathbf{n}j) + P_{31} \cos(\mathbf{n}k) \} dS = \mathbf{P}_1 d\mathbf{S}, \quad (99)$$

and so the resultant becomes

$$F = \oint (\mathbf{P}_1 - p\mathbf{i}) d\mathbf{S}. \quad . . . . . (100)$$

This integral is to be taken over the sphere of radius  $R$ . We can save much calculation, however, by noting that the divergence of  $(\mathbf{P}_1 - p\mathbf{i})$  vanishes. For

$$\text{div}(\mathbf{P}_1 - p\mathbf{i}) = \text{div} \mathbf{P}_1 - \frac{\partial p}{\partial x};$$

but this expression is zero, according to our assumptions, as may be seen from the component form of equation (83) (p. 217). Application of Gauss's theorem then shows that the integral has the same value if taken over a concentric sphere whose radius is allowed to become infinitely large. Since the surface element of this sphere increases as the second power of  $r$ , we can drop all terms of higher order than  $1/r$ , e.g. terms like  $x^2/r^5$ , in (92). For  $r$  large, the velocity components then simplify to

$$v_{x\infty} = v_0 + \frac{a}{2r} + \frac{a}{2} \frac{x^2}{r^3}, \quad v_{y\infty} = \frac{a}{2} \frac{xy}{r^3}, \quad v_{z\infty} = \frac{a}{2} \frac{xz}{r^3}. \quad (95')$$

Here  $a$  stands for  $-\frac{3}{2} v_0 R$ . In order to calculate the stresses we need the derivatives

$$\left. \begin{aligned} \left( \frac{\partial v_x}{\partial x} \right)_\infty &= \frac{a}{2} \frac{x}{r^3} \left( 1 - \frac{3x^2}{r^2} \right) \\ \left( \frac{\partial v_x}{\partial y} \right)_\infty + \left( \frac{\partial v_y}{\partial x} \right)_\infty &= - \frac{3axyx^2}{r^5} \\ \left( \frac{\partial v_x}{\partial z} \right)_\infty + \left( \frac{\partial v_z}{\partial x} \right)_\infty &= - \frac{3axzx^2}{r^5} \end{aligned} \right\} . . . . . (101)$$

Then by equations (79) and (96),

$$\begin{aligned}
 F &= \oint (\mathbf{P}_1 - p\mathbf{i}) d\mathbf{S} \\
 &= \oint \left\{ a\eta \frac{x}{r^3} \left( 1 - \frac{3x^2}{r^2} \right) \mathbf{i} - \frac{3a\eta yx^2}{r^5} \mathbf{j} - \frac{3a\eta zx^2}{r^5} \mathbf{k} - i \left( p_0 + \frac{a\eta x}{r^3} \right) \right\} d\mathbf{S} \\
 &= - \oint 3a\eta \frac{x^3}{r^5} \mathbf{i} d\mathbf{S} - \oint 3a\eta \frac{yx^2}{r^5} \mathbf{j} d\mathbf{S} - \oint 3a\eta \frac{zx^2}{r^5} \mathbf{k} d\mathbf{S} - \oint p_0 \mathbf{i} d\mathbf{S} \\
 &= -3a\eta \oint \frac{x^2}{r^5} (x\mathbf{i} + y\mathbf{j} + z\mathbf{k}) d\mathbf{S} - \oint p_0 \mathbf{i} d\mathbf{S} \\
 &= -3a\eta \oint \frac{x^2}{r^5} \mathbf{r} d\mathbf{S} - \oint p_0 \mathbf{i} d\mathbf{S}. \quad \dots \dots \dots (102)
 \end{aligned}$$

The last integral vanishes on account of the constancy of the integrand. In the first one,

$$\left. \begin{aligned} d\mathbf{S} &= r^2 \sin \theta d\theta d\phi, \\ \frac{x^2}{r^2} &= \cos^2 \theta, \end{aligned} \right\} \dots \dots \dots (103)$$

so that

$$\begin{aligned}
 F &= -3a\eta \int_0^\pi \int_0^{2\pi} \cos^2 \theta \sin \theta d\theta d\phi = + \frac{3a\eta \cdot 2\pi}{3} \left[ \cos^3 \theta \right]_0^\pi \\
 &= -4\pi a\eta. \quad \dots \dots \dots (104)
 \end{aligned}$$

Inserting the value of  $a$  as given by (94) (p. 220), we obtain the final formula

$$F = 6\pi\eta v_0 R. \quad \dots \dots \dots (105)$$

The force is thus directly proportional to the radius of the sphere and to the relative velocity of sphere and fluid. This law is of especial importance in the study of the atomic nature of electricity (cf. pp. 418 and 425). The law has been well substantiated for spheres falling with low velocities in viscous fluids

## 8. Surface Tension of Liquids.

### (a) Definition of the surface tension constant

A large number of familiar phenomena—for example, those associated with soap bubbles—indicate that the surface of a liquid

behaves like a stretched membrane, and that work must be done to increase the area of the surface. What causes this tension? In order to explain the cohesion of liquids we must assume that the molecules are subject to mutually attractive forces which decrease rapidly as the distance increases, and are therefore of little effect in the gaseous state.\* A group of molecules shielded from all external forces would thus assume the configuration in which the particles are at the minimum distance apart, i.e. a sphere. Every deformation of the sphere enlarges the surface and increases the average distance apart of the molecules, thus requiring that work be done against the attractive forces. If we pursue these ideas mathematically, and calculate the total potential energy of a quantity of liquid, we obtain a part proportional to the volume and another proportional to the surface. Since the former increases as the cube of the linear dimensions, the latter as the square, the surface phenomena will be more prominent when the volume considered is small.

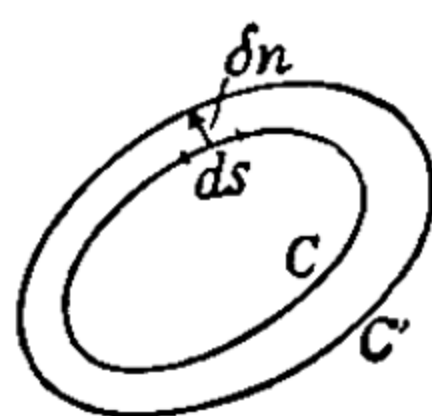


Fig. 6

We begin by considering the surface energy

$$U = aS. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (106)$$

The factor of proportionality  $a$  is called the *surface tension*. If we mark off a portion of the surface (fig. 6) by means of a closed curve  $C$ , then a force  $\tau$ , lying in the surface, acts on each unit length of this curve. Let the curve  $C$  pass into the curve  $C'$  when the surface is enlarged. Let  $\delta n$  be the displacement of an element of arc  $ds$  in the direction of the normal to  $C$  lying in the surface. Then the work done against surface tension in displacing the curve  $C$  is

$$-\delta W = \oint \tau ds \delta n. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (107)$$

Since  $\tau$  has the same value at all points of the surface, as in the case of a membrane, it may be brought outside the sign of integration. Also, since  $\oint ds \delta n$  is equal to the increase in area, we have, by (106) and (107),

$$\delta U = -\delta W = \tau \delta S = a \delta S. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (108)$$

The surface tension  $a$  is therefore equal to the force  $\tau$  acting upon unit length of a bounding curve.

\* Gravitational attraction must be assumed to exist between molecules, but it is of a much lower order of magnitude than the true molecular forces, which are electrical in nature. Repulsive forces do not come into play until the molecules are *very close* together. It is these forces which cause the molecules to occupy space.



(b) *Normal pressure at a curved surface. Differential equation of the surface*

We consider first a cylindrical surface (fig. 7) and consider the forces acting upon a surface element bounded by two generators and two curves normal to them. Since the tensions acting on these two curves lie in the surface, they are parallel to the generators, and hence can produce no resultant normal to the surface. On the other hand, the tensions acting on the generators form an angle  $d\phi$  with one another, and have components (fig. 7) directed toward the concave side of the surface. Calling the length of the

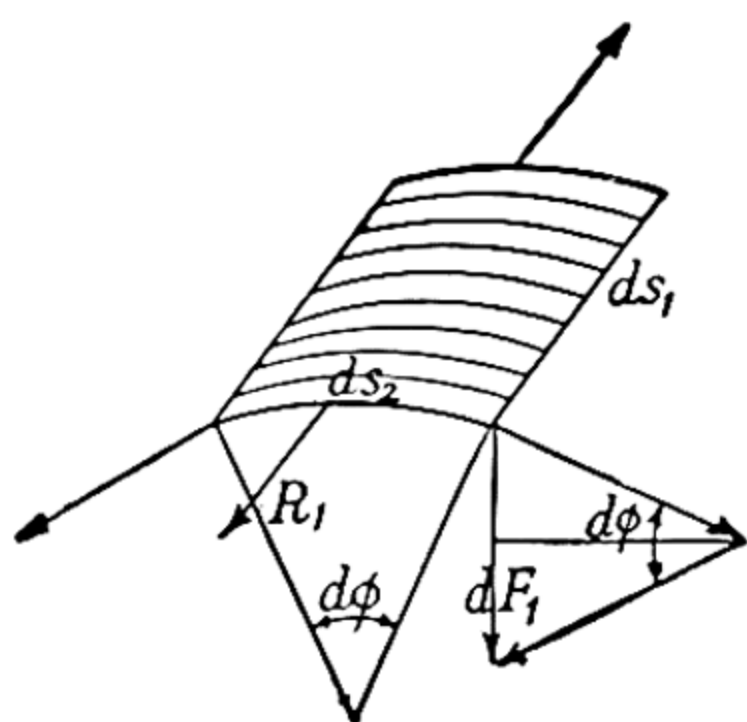


Fig. 7

generators  $ds_1$ , that of the other bounding curves  $ds_2$ , the inward component of the force acting on each generator amounts to

$$dF_1 = \tau \frac{ds_1 ds_2}{R_1} = \tau \frac{dS}{R_1} = \frac{\alpha dS}{R_1},$$

i.e.

$$p_1 = \frac{\alpha}{R_1}. \quad \dots \dots \dots (109)$$

For a surface of double curvature, we bound the surface element by two pairs of neighbouring lines of curvature. The lines of curvature are characterized by the fact that neighbouring normals to the surface, at points along these lines, intersect each other, and that the directions of their tangents are those of the normal sections of greatest and least curvature of the surface. Since the two families of lines of curvature are orthogonal, the above considerations concerning the cylinder are valid for each pair of sides of an element of a doubly-curved surface. The total normal force is then

$$dF = \alpha \left( \frac{1}{R_1} + \frac{1}{R_2} \right) dS,$$

and the inward-directed normal pressure is thus

$$p = \alpha \left( \frac{1}{R_1} + \frac{1}{R_2} \right). \quad \dots \dots \dots (110)$$

The principal radii of curvature  $R_1$  and  $R_2$  are to be considered positive if the corresponding centre of curvature lies within the fluid. The

normal pressure may assume negative values for concave surfaces, i.e. it becomes a tension.

To compute the equilibrium form of the surface we apply the principle of virtual displacements. Imagine the surface deformed in such a manner (fig. 8) that each element is given a displacement  $\delta n$  in the direction of its normal. Calling the displacement positive when directed outward, the work done against the normal pressure caused by surface tension—which is equal to the increase in potential energy—is

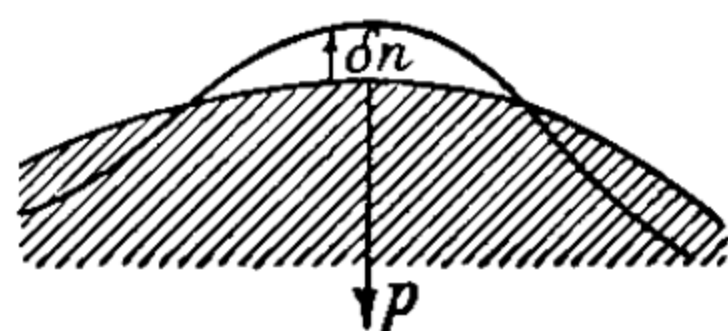


Fig. 8

$$\delta U_1 = \int_s a \left( \frac{1}{R_1} + \frac{1}{R_2} \right) dS \delta n. \quad . . . (111)$$

In addition to this change of potential energy there is another change connected with the potential of the external forces, in particular gravity; for fluid may be transferred to points of different potential as a result of the displacement. Since the energy-changes accompanying the virtual displacements are small compared with the actual differences of potential along the surface, we may, in calculating this change of potential energy, take the values for the undeformed surface, say  $\Phi$  per unit mass. Then this part of the energy change is

$$\delta U_2 = \int_s \rho \Phi dS \delta n. \quad . . . . . (112)$$

Altogether we have, for equilibrium,

$$\begin{aligned} \delta U &= \int_s a \left( \frac{1}{R_1} + \frac{1}{R_2} \right) dS \delta n + \int \rho \Phi dS \delta n \\ &= \int \left[ a \left( \frac{1}{R_1} + \frac{1}{R_2} \right) + \rho \Phi \right] dS \delta n = 0. \quad . (113) \end{aligned}$$

The displacements  $\delta n$ , however, are not entirely arbitrary. They must be so related that the volume of the liquid is not altered. But the change in volume is given by the integral  $\int dS \delta n$ , which must vanish. Hence the vanishing of the integral of equation (113) does not allow us to conclude that the integrand is zero. In order to satisfy the auxiliary condition we again apply the method of Lagrange's multipliers established on p. 115 for the case of the vanishing of the variation of a sum, and immediately applicable to the variation of an integral, which is the limit of a sum. Thus we multiply the equation

$$\delta V = \int dS \delta n = 0 \quad . . . . . (114)$$

by the undetermined multiplier  $\lambda$ , which we may write under the

integral sign, since it is constant over the surface. We then add this equation to (113). Now, from the same considerations as on p. 116, we may conclude that the vanishing of the integral implies the vanishing of the integrand, whence

$$\alpha \left( \frac{1}{R_1} + \frac{1}{R_2} \right) + \rho \Phi + \lambda = 0. \quad . \quad . \quad . \quad (115)$$

This is the differential equation of the surface of a liquid. The constant  $\lambda$  must be determined from the known values of  $R_1$ ,  $R_2$  and  $\Phi$  at some point on the surface. If, for example, we consider the rise of a liquid surface at the walls of a vessel,  $\lambda$  may be given readily for the middle point, where the surface may be considered plane. It is also convenient to reckon the gravitational potential from this point, so that  $\lambda = 0$  in this case.

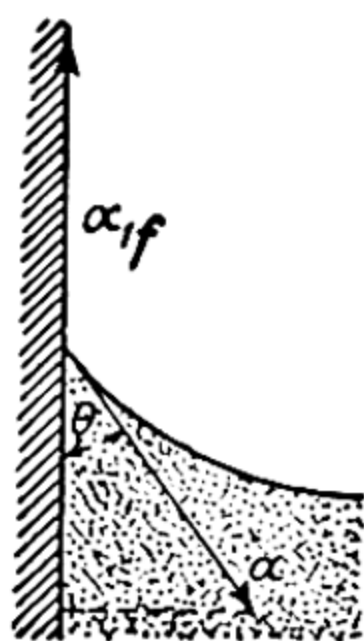


Fig. 9

### (c) Boundary conditions

In order to integrate the differential equation, the boundary conditions must also be known. The usual case is that in which the liquid is confined in a vessel whose walls intersect the surface in a boundary curve. Fig. 9 represents a section normal to the bounding curve.

At the surface of contact between the liquid and the wall there exists a surface stress  $\alpha_{1f}$ , whose magnitude depends upon the nature of the liquid and of the material composing the wall.\* For equilibrium, the component of the surface tension  $\alpha$  parallel to the wall must balance the boundary surface stress  $\alpha_{1f}$ . No condition of equilibrium need be satisfied by the normal component, for such a component manifests itself solely by a force against the wall; on the other hand, if the tangential forces were not in balance, the bounding curve would move along the wall. In the case where the liquid wets the wall, the acute angle between wall and liquid surface (fig. 9)—the so-called angle of contact—is denoted by  $\theta$ . In this case

$$\alpha \cos \theta + \alpha_{1f} = 0, \quad . \quad . \quad . \quad . \quad (116)$$

or 
$$\cos \theta = - \frac{\alpha_{1f}}{\alpha}. \quad . \quad . \quad . \quad . \quad (116')$$

This means that  $\alpha_{1f}$  is negative in this case. If

$$\left| \frac{\alpha_{1f}}{\alpha} \right| > 1,$$

\* Since the surface tension depends upon the nature of the two media, the value of the constant  $\alpha$  introduced above really is contingent upon the nature of the gas above the liquid. However, the differences for various gases are not great. Strictly speaking,  $\alpha$  is the surface tension of the liquid with respect to its own vapour.



there exists no finite angle of contact, and the liquid spreads along the wall in the form of a thin layer. This is the case for the contact between water and glass. It corresponds to the boundary condition  $\theta = 0$ . The spreading takes place even along a vertical wall. For a thin film, the work done against gravity is so small that it sets practically no limit to the spreading of the liquid.

(d) *Liquid films*

A liquid film (e.g. the wall of a soap bubble) is bounded by two parallel surfaces which are close together. Consequently, the normal pressure of a curved film is twice that corresponding to each free surface. In the interior of a spherical soap bubble of radius  $R = R_1 = R_2$ , the excess pressure is thus

$$p = \frac{4\sigma}{R}. \quad \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot (117)$$

If a bubble of this size is to be in equilibrium, the internal pressure must exceed the outside air pressure by this amount. For an unclosed film, the total pressure on both sides is the same; hence the part due to the existence of surface tension is zero, and the differential equation of an unclosed film is

$$\frac{1}{R_1} + \frac{1}{R_2} = 0. \quad \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot (118)$$

The expression on the left is called the mean curvature. Surfaces for which this quantity vanishes are called minimal surfaces, since they are the surfaces of smallest area for a given boundary curve. It is at once evident that a stretched skin, corresponding as it does to a liquid film, assumes a form having minimum area.

*Ex. 63.* Calculate the height to which a liquid will rise in a narrow tube of radius  $r_0$  whose walls are wet by the liquid ( $\cos \theta = 1$ ).

*Ex. 64.* A soap film is stretched between two equal circular wire rings lying in parallel planes, their centres being on a line normal to these planes. The resulting surface possesses rotational symmetry about this line. What is the form of the surface?

## CHAPTER X

### RELATIVISTIC MECHANICS

#### 1. Space and Time in Newtonian Mechanics.

When we spoke of the "path" and of the "velocity" of a particle in Chap. V we tacitly assumed the existence of a frame of reference (co-ordinate system) in which we could specify the position of the particle from instant to instant by giving certain numbers, and further, a means of measuring time (a clock) which marked off definite equal intervals of time at which the position could be recorded. A co-ordinate system may be given, for example, by the walls of a room, or by the positions of the stars and the direction of the plumb-line. For a measure of time, we may use any subdivision of the day, i.e. the period of rotation of the earth. Actually we find the laws of mechanics, as derived in the preceding chapters, verified when we use such reference frames and measures of time, at least to a very good approximation. But there are also frames of reference in which the fundamental law

$$\mathbf{F} = m \frac{d^2\mathbf{r}}{dt^2}$$

is *not* valid. For example, if we set our room, along with its fixed co-ordinate system, into rapid rotation, and if we allow the given applied forces to remain unchanged, then the paths of particles in the rotating co-ordinate system will be different. We can restore the validity of Newtonian mechanics by adding centripetal forces corresponding to the rotation to the other impressed forces. But what is the cause of these inertial forces? Newton gave an exact answer to this question. Despite the fact that his solution is based upon a fiction, it constitutes a foundation of the mechanics of macroscopic masses which has served for more than two centuries. Newton assumes that there exists a cosmic substratum, *absolute space*, which "in its own nature, and without reference to anything external, remains always similar and immovable". Newton continues by saying that relative space, on the other hand, is "some movable dimension or measure of the absolute space which our senses determine by its position to other bodies, and which is commonly taken for immovable space".\* Thus

\* *Principia*, Book I, Scholium to Definition VIII. English translation by Andrew Motte, New York, Daniel Adee, 1846, p. 77.



the Second Law holds only for reference frames which are at rest in absolute space or, as we shall see presently, are in uniform rectilinear motion with respect to it. The centripetal or centrifugal forces involved in rotation are caused by an acceleration of our reference frame with respect to absolute space. Thus space, which according to Kant constitutes, together with time, the perceptual form of human cognition, attains physical reality, and we can perceive its effect in the invalidity of Newtonian mechanics for accelerated frames. The newer views, on the other hand, lead to a negation of absolute space such as the ether of electrical theory. These more recent ideas recognize only relationships between parts of matter occupying space. The cause of centrifugal force is taken to be acceleration not with respect to absolute space, but with respect to the material bodies of the universe. Thus, if all the stars were removed, centrifugal force would no longer exist. Newton also makes a similar assumption concerning the second element in the form of our knowledge, viz. time: Absolute time is assumed to flow uniformly, regardless of the existence of any thinking being or any phenomenon or process suitable for measuring it. We see at once that this fiction of an absolute time also fails to stand strict examination. We need only propound the question: what happens if the velocity of all physical and chemical processes, and hence of our thought processes also, is suddenly doubled? Since we would have no means of detecting this, the absurdity underlying the very formulation of the question becomes apparent at once. Hence the concept of absolute time, which assumes that a standard clock exists somewhere in the universe, loses its justification, for it is contrary to the spirit of physics to form concepts which are neither directly nor indirectly accessible to our senses. *Nevertheless, the fact that Newton was able to formulate a thoroughly adequate foundation for the mechanics of macroscopic bodies by introducing these two concepts, whose difficulties could scarcely have escaped his notice, is to be regarded as an outstanding stroke of genius.*

## 2. Inertial Frames. The Galilean Transformation.

We start with the empirical fact that there exist reference frames for which computations based on Newton's Second Law are in complete accord with experiment. Consider, for example, celestial mechanics, which employs a stationary reference frame at the centre of gravity of the solar system. We do not now say with Newton that this frame is at rest in absolute space, or that it is in uniform rectilinear motion with respect to it (see below); we content ourselves merely with giving such a frame in which Newton's laws are valid a name. Because of the validity of the Law of Inertia, we call such a system an *inertial frame*. Evidently the confirmation of our calculations depends also upon a reasonable measurement of time. How may we obtain a cri-



terion as to whether our frame of reference, together with our clock, represents an inertial system of space and time? For this purpose we perform, at least in thought, the simplest of mechanical experiments—rectilinear motion of a particle subject to no forces. If we divide the line of motion into equal segments, we can take the time between the passing of two successive marks as the unit of time. However, *one* direction is not sufficient. If our measured path were accelerated with respect to a true inertial frame, we would obtain a non-uniform clock which would give impossible results for other experiments. It is readily seen that the necessary and sufficient condition for an inertial frame is that *three* particles projected in non-coplanar directions describe straight paths. Then, by dividing the path of any one of the particles into equal intervals, we can obtain an inertial measure of the time.

We now prove that if  $\Sigma$  is shown to be an inertial frame, Newtonian mechanics is also valid in every co-ordinate frame  $\Sigma'$  moving uniformly in a straight line with respect to  $\Sigma$ , provided that the same measure of time obtains in both systems. We shall find later that this tacit transfer of the time unit to the moving system is precisely the weak point of these almost self-evident considerations. Let the origin of  $\Sigma'$  move with the constant velocity  $u$ . Then if we denote by accents all quantities referred to the moving frame, we have

$$\mathbf{r} = \mathbf{r}' + \mathbf{u}t. \quad . . . . . (1)$$

Differentiation with respect to the time gives

$$\dot{\mathbf{r}} = \dot{\mathbf{r}}' + \mathbf{u}, \quad . . . . . (2)$$

$$\ddot{\mathbf{r}} = \ddot{\mathbf{r}}'. \quad . . . . . (3)$$

Equation (2) states that the “absolute” velocity  $\dot{\mathbf{r}}$  is equal to the vector sum of the relative velocity  $\dot{\mathbf{r}}'$  and the velocity  $\mathbf{u}$ . This theorem is general, since no use was made of the constancy of  $\mathbf{u}$  in deriving formula (2). Equation (3) states that the acceleration is the same in both systems, so that

$$\mathbf{F} = m\ddot{\mathbf{r}} = m\ddot{\mathbf{r}}', \quad . . . . . (4)$$

which was to be demonstrated. The transformation  $\mathbf{r} = \mathbf{r}' + \mathbf{u}t$  which, together with the equality of the time reckoning, corresponds to the concept of absolute space and time, is called the *Galilean transformation*. If we take the  $\dot{\mathbf{z}}$ - and the  $\dot{\mathbf{z}}'$ -axes in the direction of the velocity of motion of  $\Sigma'$ , whose magnitude is  $u$ , and if the two frames coincide at time  $t = 0$ , then the co-ordinate representation of the Galilean transformation is

$$x = x' + ut, \quad y = y', \quad z = z', \quad t = t'. \quad . . . (1')$$

The validity of the Galilean transformation carries with it the impossibility, by means of mechanical experiments, of detecting

uniform rectilinear motion with respect to absolute space; e.g. we could not determine the earth's orbital motion, which is approximately of this nature over a short interval. However, the extension of these ideas to electrodynamics (optics) shows that an experiment is possible, by means of which the reality of absolute space (the ether) might be verified. But experiments designed to test this point yielded negative results (see p. 239); this led to a revision of the previous conceptions of space and time.

### 3. Accelerated Frames of Reference. Free Fall on the Rotating Earth.

We now consider accelerated motions of the frame  $\Sigma'$  with respect to the inertial frame  $\Sigma$ . Two special cases of particular significance are selected for discussion.

#### (a) Reference frames in uniformly accelerated rectilinear motion

We assume our reference frame  $\Sigma'$  to move with uniform acceleration  $a$  in the direction of the negative  $k$ -axis. Let this frame coincide with  $\Sigma$  at the time  $t = 0$ , and let it start from rest. Such a frame is represented by a falling box. The connexion between the quantities pertaining to  $\Sigma$  and to  $\Sigma'$  is then

$$r = r' - a \frac{t^2}{2} k, \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

$$\dot{r} = \dot{r}' - atk, \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

$$\ddot{r} = \ddot{r}' - ak. \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

In  $\Sigma$  we have  $F = m\ddot{r} = m\ddot{r}' - mak$ ; hence for the same impressed forces in  $\Sigma'$ ,

$$m\ddot{r}' = F + mak. \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

In addition to the resultant of the *impressed* forces we now have an *apparent force* or *inertial force*  $mak$ . If  $a$  is identical with the acceleration of gravity, the inertial force exactly balances the gravity force  $-mgk$ . Hence the force of gravity does not operate in the falling chamber. On the other hand, if we imagine the chamber to be at a place where there is no gravitational force, and if we move it with uniform acceleration  $g$  in the direction of the  $+k$ -axis, then an apparent force of amount  $mg$ , directed along the negative  $k$ -axis, acts upon every body within the box. Thus an observer within the box would say that he is in the gravitational field of the earth. This equivalence of an accelerated frame of reference and a gravitational field is of fundamental importance in the general theory of relativity (cf. p. 258). The equivalence can be maintained, however, only if the ratio of inertial mass, i.e. the mass entering in Newton's Second Law, and gravi-

tational mass, i.e. the mass occurring in the law of gravitation, is the same for all bodies. In this case the ratio may be set equal to unity; this determines the numerical value of the gravitational constant.

Following this preliminary glimpse we turn to the other important special case.

(b) *Uniformly rotating frames of reference*

In the previously considered types of motion the unit vectors  $i'$ ,  $j'$ ,  $k'$  maintained their directions, and we assumed these directions to coincide with those of  $i$ ,  $j$ ,  $k$  respectively. For a rotational motion of  $\Sigma'$ , however, the directions of these vectors change. If we suppose the origins of the two frames to coincide, and if we take the axis of rotation through the common origin, we have

$$\mathbf{r} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k} = x'\mathbf{i}' + y'\mathbf{j}' + z'\mathbf{k}'. \quad . \quad . \quad (9)$$

In taking time derivatives, the unit vectors will thus be variable:

$$\dot{\mathbf{r}} = \frac{dx'}{dt} \mathbf{i}' + \frac{dy'}{dt} \mathbf{j}' + \frac{dz'}{dt} \mathbf{k}' + x' \frac{d\mathbf{i}'}{dt} + y' \frac{d\mathbf{j}'}{dt} + z' \frac{d\mathbf{k}'}{dt}. \quad (10)$$

The first three terms represent the velocity relative to  $\Sigma'$ , the last three give the velocity of a point rigidly attached to  $\Sigma'$ . If we denote the angular velocity vector by  $\boldsymbol{\omega}$ , then

$$\frac{d\mathbf{i}'}{dt} = [\boldsymbol{\omega}\mathbf{i}'], \quad \frac{d\mathbf{j}'}{dt} = [\boldsymbol{\omega}\mathbf{j}'], \quad \frac{d\mathbf{k}'}{dt} = [\boldsymbol{\omega}\mathbf{k}']. \quad . \quad (11)$$

If these values are substituted in equation (10), the last three terms may be combined:

$$x'[\boldsymbol{\omega}\mathbf{i}'] + y'[\boldsymbol{\omega}\mathbf{j}'] + z'[\boldsymbol{\omega}\mathbf{k}'] = [\boldsymbol{\omega}(x'\mathbf{i}' + y'\mathbf{j}' + z'\mathbf{k}')] = [\boldsymbol{\omega}\mathbf{r}]. \quad (12)$$

If we denote differentiation referred to  $\Sigma'$  by  $d'/dt$ , then (10) may be written

$$\frac{d\mathbf{r}}{dt} = \frac{d'\mathbf{r}}{dt} + [\boldsymbol{\omega}\mathbf{r}]. \quad . \quad . \quad . \quad . \quad . \quad (13)$$

This is true for any vector, since we can insert the components of any vector whatsoever in place of  $x$ ,  $y$ ,  $z$ . Symbolically, we may thus write

$$\frac{d}{dt} = \frac{d'}{dt} + [\boldsymbol{\omega}]. \quad . \quad . \quad . \quad . \quad . \quad (14)$$

Differentiating (13) again with respect to  $t$  and applying the above rule, we have, remembering that  $\boldsymbol{\omega}$  is constant,

$$\begin{aligned} \frac{d^2\mathbf{r}}{dt^2} &= \frac{d'}{dt} \left( \frac{d\mathbf{r}}{dt} \right) + \left[ \boldsymbol{\omega} \frac{d\mathbf{r}}{dt} \right] \\ &= \frac{d'}{dt} \left( \frac{d'\mathbf{r}}{dt} + [\boldsymbol{\omega}\mathbf{r}] \right) + \left[ \boldsymbol{\omega} \left( \frac{d'\mathbf{r}}{dt} + [\boldsymbol{\omega}\mathbf{r}] \right) \right], \quad (15) \end{aligned}$$



$$\text{or} \quad \frac{d^2 \mathbf{r}}{dt^2} = \frac{d^2 \mathbf{r}'}{dt'^2} + 2 \left[ \boldsymbol{\omega} \frac{d' \mathbf{r}}{dt} \right] + [\boldsymbol{\omega} [\boldsymbol{\omega} \mathbf{r}]]. \quad (16)$$

The first term represents the acceleration relative to  $\Sigma'$ , the third represents the acceleration in the system  $\Sigma$  of a point fixed to  $\Sigma'$ . However, it is no longer true, as in the case of velocity, that the absolute acceleration is equal to the sum of these two terms; there is, in addition, a third term—twice the vector product of angular velocity by relative velocity. This term, which occurs only when the particle moves in  $\Sigma'$ , is called the *Coriolis acceleration*, after its discoverer. The Second Law then becomes

$$\mathbf{F} = m \frac{d^2 \mathbf{r}}{dt^2} = m \frac{d^2 \mathbf{r}'}{dt'^2} + 2m \left[ \boldsymbol{\omega} \frac{d' \mathbf{r}}{dt} \right] + m [\boldsymbol{\omega} [\boldsymbol{\omega} \mathbf{r}]], \quad (17)$$

or, in the moving system,

$$m \frac{d^2 \mathbf{r}'}{dt'^2} = \mathbf{F} - 2m \left[ \boldsymbol{\omega} \frac{d' \mathbf{r}}{dt} \right] - m [\boldsymbol{\omega} [\boldsymbol{\omega} \mathbf{r}]]; \quad (17')$$

that is, we have, in addition to the impressed forces, the ordinary centrifugal force  $-m[\boldsymbol{\omega} [\boldsymbol{\omega} \mathbf{r}]]$  and the Coriolis force

$$-2m \left[ \boldsymbol{\omega} \frac{d' \mathbf{r}}{dt} \right].$$

As an example of the forces arising from rotation of the frame of reference we take the case of free fall on the rotating earth. Co-ordinate frames rigidly attached to the earth are, strictly speaking, not inertial frames, on account of the daily rotation. The angular velocity of the orbital motion, on the other hand, can be neglected, having to the other the ratio 1 : 365. However, the deviation from a true inertial frame is so small that it usually may be neglected. Nevertheless, there are various mechanical experiments which permit the rotation of the earth to be demonstrated. These experiments were formerly interpreted as proof of the rotation with respect to absolute space; from the point of view of the present day, however, they may be considered merely as evidence of rotation with respect to the other matter of the universe. Take an inertial frame whose  $k$ -axis is in the direction of the earth's axis. Let the frame  $\Sigma'$  be oriented as follows: Take the  $k'$ -axis in the direction of the plumb-line at the geographical latitude  $\phi$ . This direction is that of the resultant of gravity and centrifugal force. This direction will not pass through the centre of the earth if we consider the earth to be spherical, but it is sufficiently accurate to take the angle between this direction and the earth's axis to be

$$\left( \frac{\pi}{2} - \phi \right).$$

We also omit from consideration the polar flattening of the earth. The positive sense is taken to be that pointing outward from the earth's centre. The  $\hat{z}'$ -axis is directed southward, the  $\hat{j}'$ -axis then points eastward (fig. 1). We then have the equation of motion

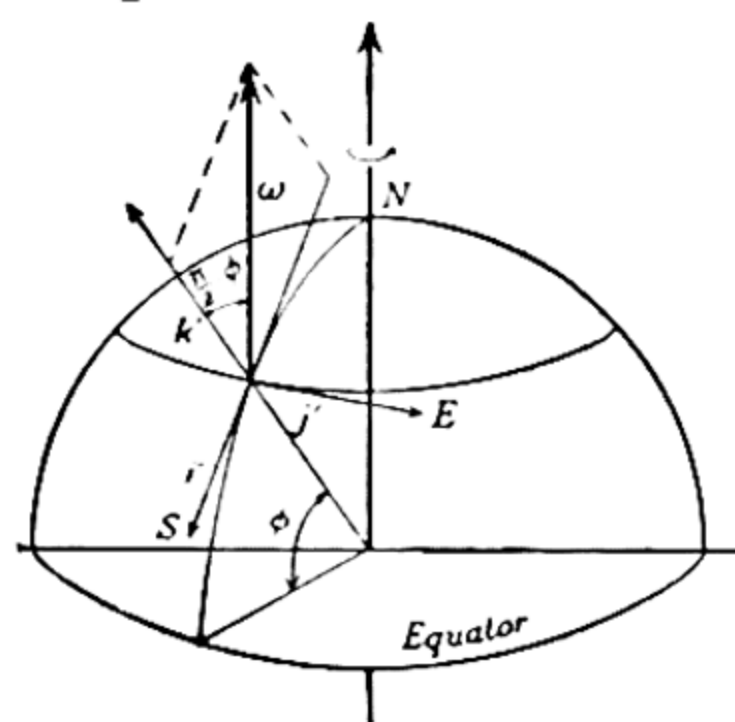


Fig. 1

$$m \frac{d^2 \mathbf{r}}{dt^2} = -mg\mathbf{k}' - 2m \left[ \boldsymbol{\omega} \frac{d\mathbf{r}}{dt} \right]. \quad (18)$$

The centrifugal force is already included in the term  $-mg\mathbf{k}'$ , since we understand  $g$  to represent the resultant acceleration of a falling body at the point considered.

The angular velocity  $\boldsymbol{\omega}$  has the components  $-\omega\hat{z}' \cos \phi$  and  $+\omega\hat{k}' \sin \phi$  in the frame  $\Sigma'$ . In co-ordinates we thus have

$$\left. \begin{aligned} m \frac{d^2 x'}{dt^2} &= 2m\omega \sin \phi \frac{dy'}{dt} \\ m \frac{d^2 y'}{dt^2} &= -2m\omega \sin \phi \frac{dx'}{dt} - 2m\omega \cos \phi \frac{dz'}{dt} \\ m \frac{d^2 z'}{dt^2} &= -mg + 2m\omega \cos \phi \frac{dy'}{dt} \end{aligned} \right\} \quad (18')$$

Leaving the rotation of the earth out of consideration, the  $\hat{z}'$ - and  $\hat{j}'$ -components of velocity vanish for free fall. Hence, even when the rotation is taken into account, the quantities  $dx'/dt$  and  $dy'/dt$  are small compared with  $dz'/dt$  so that we have the simplified equations

$$\left. \begin{aligned} \frac{d^2 x'}{dt^2} &= 0 \\ \frac{d^2 y'}{dt^2} &= -2\omega \cos \phi \frac{dz'}{dt} \\ \frac{d^2 z'}{dt^2} &= -g \end{aligned} \right\} \quad (19)$$

The first of these equations shows that no deviation occurs in the north-and-south direction. Integrating the third equation and taking

$$z'_0 = \left( \frac{dz'}{dt} \right)_0 = 0,$$

there results 
$$\frac{dz'}{dt} = -gt, \quad z' = -\frac{g}{2}t^2. \quad . . . . . (20)$$

Inserting this value in the second equation,

$$\frac{d^2y'}{dt^2} = +2\omega gt \cos \phi. \quad . . . . . (21)$$

Integrating twice and taking

$$y'_0 = \left(\frac{dy'}{dt}\right)_0 = 0,$$

we obtain 
$$y' = \frac{1}{3}\omega gt^3 \cos \phi. \quad . . . . . (22)$$

This result—an eastward deviation of falling bodies, proportional to the cube of the time of fall—agrees with observation.

*Ex. 65. The Foucault pendulum.*—Calculate the motion of a freely swinging pendulum bob suspended by a thread, taking into account the motion of the earth. (The  $z$ -component of the velocity may be neglected here.)

*Ex. 66.* Using equation (14), express the total change in angular momentum of a rigid body in terms of one part relative to the body and another due to rotation of the system of co-ordinates fixed in the body. Using the result, write the equation  $\mathbf{M} = d\mathbf{P}/dt$  in terms of components of  $\omega$  resolved along the principal axes of inertia. These equations are called Euler's equations.

#### 4. Moving Frames of Reference in Acoustics. The Doppler Effect.

Before discussing the difficulties which the concept of absolute space raises in optics, we shall consider the analogous questions for sound. Here the relations involved are perfectly clear and the effects are quite definite, on account of the fact that the velocity of sound

in air is not too great compared with mechanically attainable speeds. We know that the velocity of propagation of sound in air is  $c = 333$  m./sec. In a co-ordinate system fixed with respect to the source of sound (taken to be at rest) and with respect to the air, the sound will have arrived at a sphere of radius  $c$  after one second.

Let us now lay out a measuring rod and move it relative to the air. We then measure the relative velocity  $c_r$  along the moving measuring rod, the direction of  $c_r$  being given by that of the rod. Now the absolute velocity must always be  $c$ ; hence if the velocity of the measuring rod is given, in direction and magnitude, by  $v$ ,

$$c = c_r + v \quad \text{or} \quad c_r = c - v. \quad . . . . . (23)$$

In particular, if the rod is moved along its own length (position  $b$ , fig. 2),

$$c_r = c - v; \quad . . . . . (23')$$

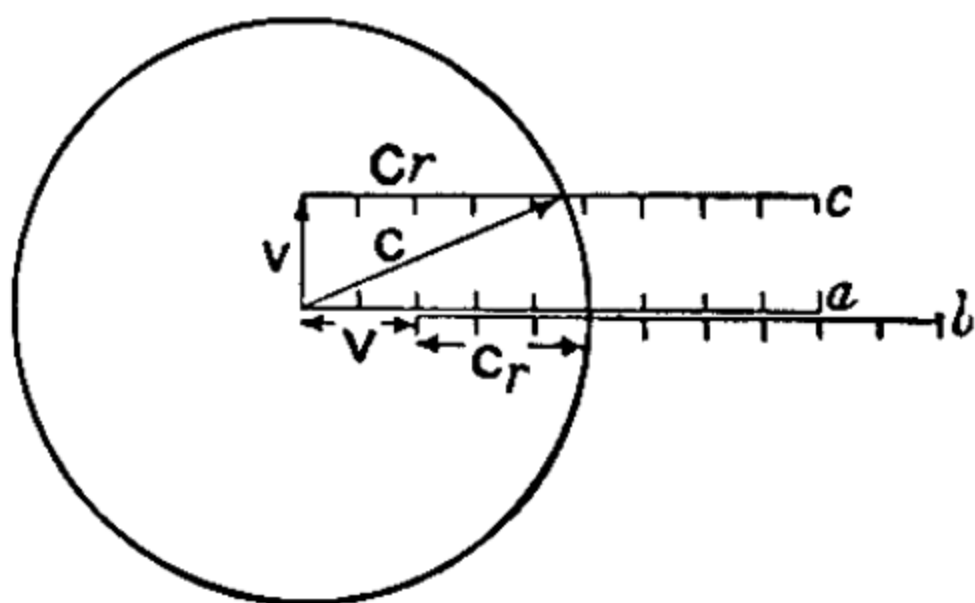


Fig. 2



for motion normal to its length (position  $c$ , fig. 2),

$$c_r = \sqrt{c^2 - v^2}. \quad . \quad . \quad . \quad . \quad . \quad (23'')$$

Moreover, we know that motion relative to the source of sound is accompanied by a change in pitch of the tone perceived (Doppler effect). First imagine the source and the observer both at rest. The source emits  $\nu_s$  condensations per second; these follow each other at distances apart  $\lambda_0 = c/\nu_s$ . The succession of sound waves may be thought of as a long rod moving forward with the speed  $c$ , the condensations being located on it, like a series of beads, with a separation equal to  $\lambda_0$ . Hence  $\nu_M = c/\lambda_0 = \nu_s$  is the number of condensations passing the position of the observer each second. If we now permit the observer to move toward the source of sound, then he receives an additional number of vibrations per second equal to the number of waves in the distance he covers per second, i.e. a number  $v/\lambda_0$ . Hence

$$\nu_M = \nu_s + \frac{v}{\lambda_0} = \nu_s + \nu_s \frac{v}{c} = \nu_s \left( 1 + \frac{v}{c} \right). \quad . \quad . \quad (24)$$

For motion away from the source the positive sign in the parenthesis is to be replaced by the negative.

The Doppler effect due to the *motion of the source* is essentially different, although the effect is in the same direction, i.e. the pitch is raised when the distance between source and observer is decreasing, and lowered when the distance is increasing. The situation may be analysed in this way: If the source were at rest, then the first disturbance—say a condensation—would be at a distance  $\lambda_0$  from the source after the time of one vibration had elapsed. But the source has moved on a distance  $v/\nu = (v/c)\lambda_0$  in this time, thus decreasing the distance between the source and the first condensation by  $(v/c)\lambda_0$ . This decrease holds for all distances between condensations; for after another vibration, the first disturbance has progressed a distance  $\lambda_0$ , but the second starts from the new position of the source, and so the distance between the first and second condensations is again decreased by  $(v/c)\lambda_0$ , and so on. Fig. 3 shows how this occurs. The number of vibrations passing the observer each second is thus

$$\frac{c}{\lambda'} = \frac{c}{\lambda_0(1 - v/c)},$$

or

$$\nu_M = \frac{\nu_s}{1 - v/c}. \quad . \quad . \quad . \quad . \quad . \quad (25)$$

If the source moves away from the observer, the positive sign must be used before the second term in the denominator. Comparing this result with the case where the observer alone was in motion, we see that

the two are equivalent only to a first approximation, i.e. when higher powers of  $v/c$  can be neglected. The reason is that the medium carrying

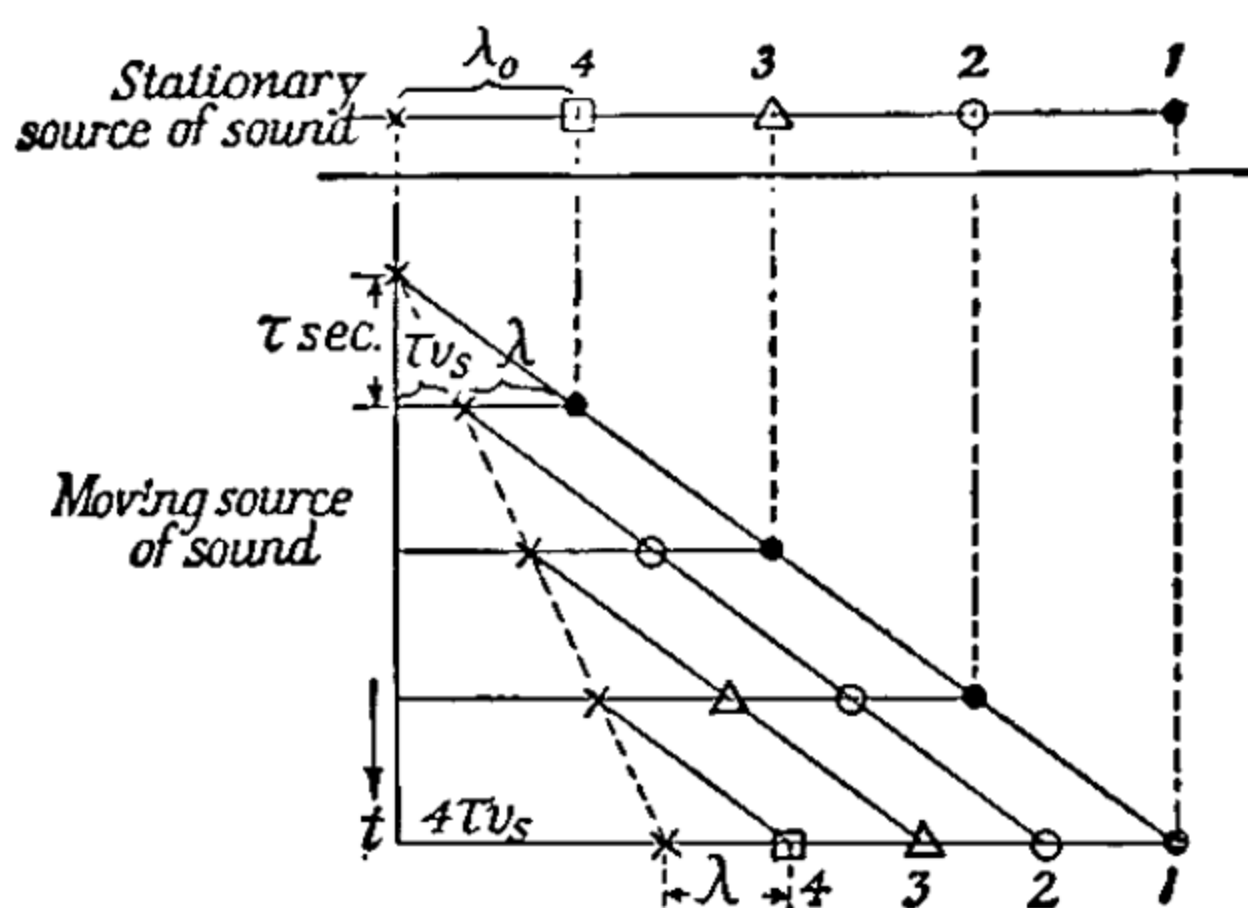


Fig. 3

the sound (the air) provides a special co-ordinate frame  $\Sigma$  to which all motion is to be referred. One sees at once that the two cases are essentially different by taking  $v = c$ . In this instance we have  $\nu_M = 2\nu_B$  for motion of the observer, while if the source is moving,  $\nu_M = \infty$ .

### 5. Moving Frames of Reference in Optics. The Michelson-Morley Experiment.

According to the older ideas, absolute space—called the ether in electrical theory—is the carrier of electromagnetic and therefore of optical phenomena. It is therefore natural to apply the results obtained for sound to optics. Actually, however, this procedure leads to unsurmountable difficulties. The existence of an optical Doppler effect of approximately the magnitude given by equation (25) has been observed in the spectra of stars moving toward or away from the earth, as well as in the light from canal (or positive) rays—fast-moving luminous atoms (cf. p. 441). If we could increase the velocity of the light source, or of the observer, or increase the accuracy of measurement to the extent that quantities of the second order in  $v/c$  could be measured, then we would be able to test the existence of the ether. As we saw in Acoustics, if we have a medium at rest, the second-order quantities show that the Doppler shift depends not only upon the relative velocity of source and observer, but on the velocity with respect to the medium. Unfortunately, there is little hope of securing an experimental decision concerning the reality of the ether concept in this way. Only recently has it been found possible to increase the accuracy of Doppler-effect measurements on positive rays

to the extent that second-order quantities can be measured (see p. 246). It is easier to carry out, with sufficient accuracy, an experiment by which the velocity of propagation with respect to a frame moving through the ether is investigated. Since light is propagated in the ether with velocity  $c$ , the velocity of light relative to a system moving through the ether with velocity  $v$  is  $c - v$ . A direct measurement or comparison of the velocity of light in various directions on the earth, which (considering only the orbit about the sun) moves uniformly with a velocity of about 30 km./sec., is again, it is true, beyond experimental possibility. A comparison is made possible, however, by using the light itself, in a somewhat different arrangement, to measure the

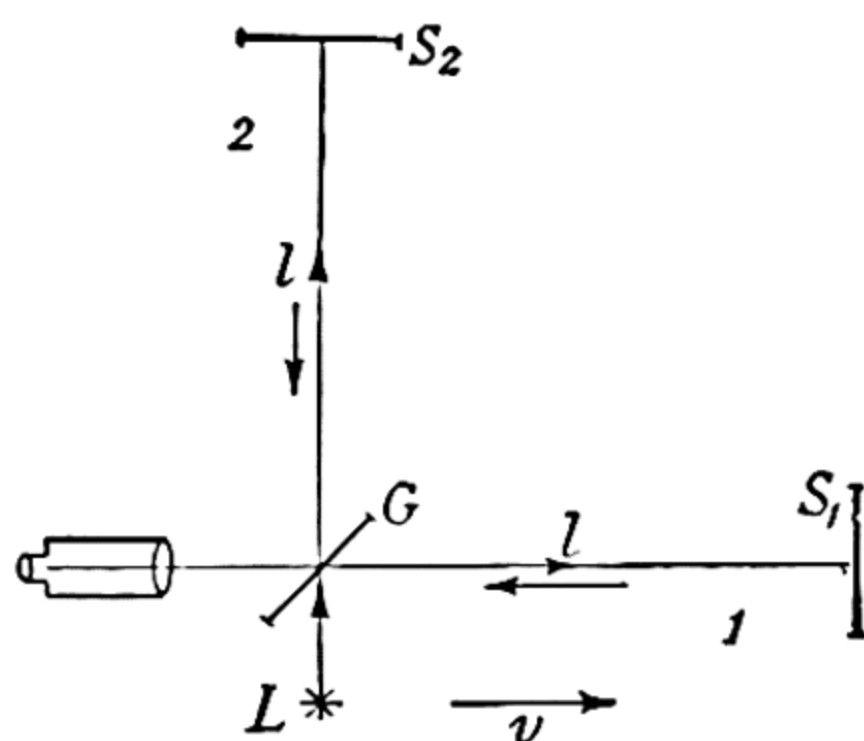


Fig. 4

vanishingly small time differences. This scheme, which had already been suggested by Maxwell, was first utilized by Michelson and Morley in 1880. It has since been tested many times by various investigators. In this arrangement (fig. 4) light from the source  $L$  is divided into two beams, normal to each other, by the half-silvered plane parallel glass plate  $G$ . Each beam covers the distance  $l$  to the mirrors  $S_1$  or  $S_2$ , and is reflected back over its original path. Assume the arm 1 to be in the direction of the earth's motion; then the relative velocity of the light is  $c - v$  on the initial path and  $c + v$  on the return. Hence the time taken to pass to and fro along this arm is

$$t_1 = \frac{l}{c - v} + \frac{l}{c + v} = \frac{2l}{c} \left( 1 + \frac{v^2}{c^2} + \dots \right). \quad \dots (26)$$

In the direction normal to this path the relative velocity of light is  $\sqrt{c^2 - v^2}$ ; hence the time required for the double journey along the arm 2 is

$$t_2 = \frac{2l}{\sqrt{c^2 - v^2}} = \frac{2l}{c} \left( 1 + \frac{1}{2} \frac{v^2}{c^2} \right). \quad \dots (27)$$



Thus there exists a difference—a very small one—between the two times:

$$\Delta t = t_1 - t_2 = \frac{l}{c} \cdot \frac{v^2}{c^2} \quad . \quad . \quad . \quad . \quad . \quad (28)$$

If we take the orbital speed of the earth to be 30 km./sec. and let the path  $l$  be 30 m. long,  $\Delta t = 10^{-15}$  sec. In spite of its smallness, this difference can be detected if we use the light vibrations themselves instead of a material clock for this measurement. Now the number of vibrations  $N$  described by yellow light ( $\lambda = 6 \times 10^{-5}$  cm.) in  $10^{-15}$  sec. is

$$N = \frac{\Delta t}{T} = \Delta t \cdot \nu = \Delta t \cdot \frac{c}{\lambda} = 10^{-15} \cdot \frac{3 \cdot 10^{10}}{6 \cdot 10^{-5}} = \frac{1}{2} \quad . \quad (29)$$

The difference is thus one half of one vibration. If the apparatus is turned through  $90^\circ$  the two arms interchange their rôles, and the difference has the opposite sign, so that the entire difference for two positions  $90^\circ$  apart amounts to one complete vibration. Since both partial beams originate in the same light source, they interfere when reunited (see p. 410) and we obtain a system of interference fringes which may be thought of as "curves of equal thickness" in the wedge formed by  $S_1$  and the image of  $S_2$  produced by reflection in the semi-silvered plate. Starting with the arm 1 in the direction of the motion of the earth, we have a definite system of fringes. If, now, the apparatus is turned through  $45^\circ$ , the two arms are symmetrically disposed, and the original time difference of  $T/2$ , corresponding to a phase difference of  $\pi$ , vanishes. The places which were originally dark are now bright, i.e. the fringe system has been displaced by half a fringe width. If we turn the apparatus further in the same sense, the opposite difference obtains, and when the rotation amounts to  $90^\circ$  the displacement is an entire fringe width. This is, of course, subject to the assumption of the numerical values adopted above. The result of the Michelson-Morley experiment, conducted under approximately these conditions, was the absence of any such fringe displacement. When, in 1925, D. C. Miller reported a positive effect corresponding to a velocity of 10 km./sec., verifications were attempted by various investigators. Joos, working with improved apparatus at the Zeiss plant in Jena, was able to show that if a positive effect does exist, it must be smaller than one thousandth of the width of a fringe. Fig. 1, Plate I, shows a record of the interference fringes. This record was obtained by moving a photographic plate in front of a 0.2 mm.-wide slit mounted normal to the fringes, the entire apparatus being rotated, meanwhile. Instead of a sinuous line repeating after one half turn, we see that the stripes are perfectly straight. *Experiment has thus decided against our acoustic analogy and against the existence of a stationary medium carrying light; i.e. the existence of a cosmic ether or absolute space is disproved.*

## 6. The Relativistic Conception of Space and Time. The Lorentz Transformation.

Disregarding hypotheses which are contrived *ad hoc*, and which are therefore unsatisfactory, the result of the Michelson-Morley experiment leads compulsorily to a revision of our concepts concerning space and time. The negative result of this experiment makes it very evident that absolute space is a fiction. But absolute time must also be abandoned. The definitions occurring in the treatment of time by the old mechanics are imperfect. For example, what is the meaning of the statement: "Two events took place simultaneously at  $A$  and  $B$ "? If we postulate an absolute, uniformly-flowing time, there can be no doubt as to the meaning of this statement; but physics demands an experimental criterion of the correctness of this statement, and this is where the difficulties begin. We must have synchronized clocks at the stations  $A$  and  $B$  which show the absolute time. Then the events may be said to be simultaneous if the clocks indicate the same time at the instant when the events occur. But how can the clocks be synchronized? An arrangement which suggests itself is the following: Place two mirrors, inclined at  $45^\circ$  to the path  $AB$ , at the mid-point  $C$  of  $AB$ , so that an observer at  $C$  can see the two stations  $A$  and  $B$  simultaneously. We shall say that the two clocks are synchronized if they indicate the same time at every instant, when seen at  $C$ . But from the standpoint of absolute concepts, they will not be exactly synchronous even in this case, for the experiment is performed on the earth, which moves relative to the ether; hence the light by which we see the clocks will require a time  $l/(c - v)$  to come to us from one clock, and a time  $l/(c + v)$  to come from the other, assuming the path  $AB$  (of length  $2l$ ) to be in the direction of the earth's motion. What value is to be put for  $v$ ? There is certainly a velocity of the solar system relative to the fixed stars to be added to the orbital velocity of the earth, and probably a very large velocity of the galaxy with respect to the extra-galactic systems. From the absolute point of view, however, there may also be a common motion of all matter with respect to the ether. The result of the Michelson-Morley experiment relieves us of the necessity of answering this question, for it shows that the velocity of light is the same in all directions on the moving earth. Therefore the above experimental criterion for the synchronization of the clocks of a system remains intact. But, as was first recognized by H. A. Lorentz, the result of the Michelson-Morley experiment requires that two systems in relative motion have two different time measures, for it is observed in each system that light reaches a sphere of radius  $c$  after one second. Hence the Galilean transformation, which sets  $t' = t$ , cannot be correct. Its insufficiency was not recognized sooner on account of the fact that the velocity of light is so great in com-



parison with all mechanical velocities that the deviations can be detected only by the most refined optical measurements. The fact that the Galilean transformation is incompatible with the result of the Michelson-Morley experiment may be seen by applying the transformation to a spherical wave observed in a system at rest. The result is the equation of a sphere eccentric with respect to the origin. We must therefore transform  $t$ , along with the other quantities, in such a way that the transformation formulæ give a spherical wave for the propagation of light in every frame of reference, i.e. the identity

$$x^2 + y^2 + z^2 - c^2t^2 \equiv x'^2 + y'^2 + z'^2 - c^2t'^2 = 0 \quad . \quad (30)$$

must be satisfied.

For the sake of simplicity we take the  $x$ -axis in the direction of the relative velocity of  $\Sigma$  and  $\Sigma'$ , thus permitting the  $x$ -axis to coincide with the  $x'$ -axis. We then make the obvious assumption that the transverse co-ordinates remain unchanged. Thus the identity reduces to

$$x^2 - c^2t^2 \equiv x'^2 - c^2t'^2 = 0. \quad . \quad (30')$$

Since there is but *one* relative velocity  $v$ , the transformation formulæ must further fulfil the condition that  $O'$  has the co-ordinate  $vt$  in  $\Sigma$ , and that  $O$  has the co-ordinate  $-vt'$  in  $\Sigma'$  (fig. 6). The simplest relations which accomplish this end are

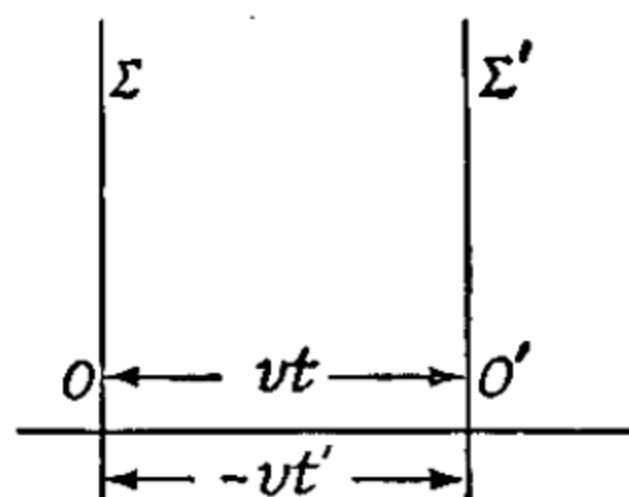


Fig. 6

$$\left. \begin{aligned} x' &= k(x - vt) \\ x &= k'(x' + vt') \end{aligned} \right\} \quad . \quad . \quad . \quad . \quad (31)$$

and

We eliminate  $x'$  from the second equation by inserting its value obtained from the first, and find

$$t' = k \left[ t - \frac{x}{v} \left( 1 - \frac{1}{kk'} \right) \right]. \quad . \quad . \quad . \quad . \quad (32)$$

Putting the values of  $x'$  and  $t'$  in the right member of (30'), we have

$$\begin{aligned} x^2 - c^2t^2 - k^2(x^2 - 2xvt + v^2t^2) \\ + c^2k^2 \left[ t^2 - \frac{2xt}{v} \left( 1 - \frac{1}{kk'} \right) + \frac{x^2}{v^2} \left( 1 - \frac{1}{kk'} \right)^2 \right] \equiv 0. \end{aligned} \quad (33)$$

The identity requires that the coefficients of  $x^2$ ,  $xt$  and  $t^2$  vanish separately; to accomplish this, we have only the two constants  $k$  and  $k'$  at our disposal. The coefficient of  $x^2$  yields

$$1 - k^2 + \frac{c^2}{v^2}k^2 - 2\frac{c^2}{v^2}\frac{k}{k'} + \frac{c^2}{v^2} \cdot \frac{1}{k'^2} = 0.$$



The coefficient of  $t^2$  requires that

$$-c^2 - k^2 v^2 + k^2 c^2 = 0 \quad \text{or} \quad k = \frac{1}{\sqrt{1 - v^2/c^2}}.$$

The coefficient of  $xt$  gives

$$(v^2 - c^2)kk' + c^2 = 0,$$

or, after inserting the value of  $k$ ,

$$k' = \frac{1}{\sqrt{1 - v^2/c^2}}.$$

It is now easy to verify that these values of  $k$  and  $k'$  make the coefficient of  $x^2$  vanish also.

We thus have the formulæ of transformation

$$\left. \begin{aligned} x' &= \frac{x - vt}{\sqrt{1 - \frac{v^2}{c^2}}} & x &= \frac{x' + vt'}{\sqrt{1 - \frac{v^2}{c^2}}} \\ t' &= \frac{t - \frac{v}{c^2}x}{\sqrt{1 - \frac{v^2}{c^2}}} & t &= \frac{t' + \frac{v}{c^2}x'}{\sqrt{1 - \frac{v^2}{c^2}}} \end{aligned} \right\} \dots \dots (34)$$

If we let  $c$  tend to infinity we obtain in the limit the Galilean transformation, which is a first approximation to the Lorentz transformation for finite  $v$ . The ratio  $v/c$  will, as is customary, be denoted by  $\beta$  in what follows. Using (34), we readily see by differentiation that the speed  $dx/dt$  of  $O'$  with respect to  $\Sigma$ , and the speed  $dx'/dt'$  of  $O$  with respect to  $\Sigma'$ , are both equal to  $v$ .

## 7. Immediate Consequences of the Lorentz Transformation.

### (a) *Relativity of simultaneity and of the order of events*

Let two events occur simultaneously at the points  $x_1$  and  $x_2$  in the frame  $\Sigma$ , i.e. let  $t_1 = t_2$ . Then

$$t_1' = \frac{t_1 - \frac{v}{c^2}x_1}{\sqrt{1 - \beta^2}}, \quad t_2' = \frac{t_2 - \frac{v}{c^2}x_2}{\sqrt{1 - \beta^2}}, \quad \dots \dots (35)$$

i.e.  $t_1'$  and  $t_2'$  are different. Two events at different places  $x_1$  and  $x_2$  which are simultaneous for an observer at rest in the frame  $\Sigma$  no longer appear so to an observer moving relative to  $\Sigma$ . We ask further: "May even the order of two events depend upon the state of

motion of the observer?" Let us take  $t_2 > t_1$ , i.e.  $t_2 - t_1 > 0$ . Then

$$t_2' - t_1' = \frac{t_2 - t_1 - \frac{v}{c^2}(x_2 - x_1)}{\sqrt{1 - \beta^2}}. \quad \dots \quad (35')$$

In order to obtain a real time difference  $t_2' - t_1'$  at all in the accented system, the velocity  $v$  of  $\Sigma'$  must be less than  $c$ ;  $t_2' - t_1'$  certainly remains positive as long as

$$t_2 - t_1 > \frac{x_2 - x_1}{c}.$$

Now we shall see below that the greatest velocity with which an agency connecting a cause and an effect can be propagated is  $c$ . If there were a more rapid process, we would use it in place of light to define simultaneity. Hence, as long as  $c$  is the greatest velocity attainable, the order of two causally connected events cannot be reversed. However, the order of two events which are not causally connected may be reversed consistently with the above relationship.

### (b) *The Einstein time dilatation*

Let a clock at  $x_1$  give signals at intervals

$$\Delta t = t_2 - t_1.$$

Seen from the moving system, the interval will be, by equation (34),

$$\Delta t' = \frac{t_2 - t_1}{\sqrt{1 - \beta^2}} = \frac{\Delta t}{\sqrt{1 - \beta^2}}. \quad \dots \quad (36)$$

Thus the interval  $\Delta t$  appears to the moving observer to be lengthened.

### (c) *The Lorentz contraction*

We perform a measurement of length by laying a measuring rod alongside the segment to be measured and reading the interval between the division marks which coincide *at a given instant* with the ends of the length to be measured. As long as the measuring rod and the interval to be measured are at rest with respect to one another, this definition is trivial; not so, however, if there exists a relative velocity  $v$  between the two. Let the distance  $l = x_2 - x_1$  between the ends of the interval be measured by means of a linear scale at rest in this frame. According to equations (34) (p. 242), a moving observer will note

$$x_2' - x_1' = \frac{x_2 - x_1 - v(t_2 - t_1)}{\sqrt{1 - \beta^2}}. \quad \dots \quad (37)$$

The times  $t_2$  and  $t_1$ , however, are to be determined in such a way that coincidence with the scale divisions occurs at the same instant *in the observer's system*, i.e. we must have, not  $t_2 = t_1$ , but  $t_2' = t_1'$ . From the Lorentz transformation we have, for  $t_2' = t_1'$ ,

$$t_2 - t_1 = (x_2 - x_1) \frac{v}{c^2},$$

and so 
$$l' = x_2' - x_1' = (x_2 - x_1) \sqrt{1 - \beta^2}. \quad . \quad . \quad (38)$$

Seen by the moving observer, linear dimensions are reduced in the ratio  $\sqrt{1 - \beta^2} : 1$ . A body which appears to be spherical to an observer at rest relative to it will thus appear to a moving observer to be an oblate spheroid.

(d) *Einstein's velocity addition theorem*

In Newtonian mechanics, relative velocities are found by simple addition or subtraction (cf. § 2, p. 229). The composition of velocities in a mechanics based on the Lorentz transformation is more involved. Let a frame  $\Sigma'$  move relative to  $\Sigma$  in the direction of the  $x$ -axis with velocity  $v$ . Let a point in  $\Sigma'$  have the velocity

$$\mathbf{u}' = u_x' \mathbf{i}' + u_y' \mathbf{j}' + u_z' \mathbf{k}',$$

where 
$$u_x' = \frac{dx'}{dt'}, \quad u_y' = \frac{dy'}{dt'}, \quad u_z' = \frac{dz'}{dt'}. \quad . \quad . \quad (39)$$

What are its velocity components  $u_x, u_y, u_z$ , in  $\Sigma$ ? It follows from the Lorentz transformation, by differentiation, that

$$u_x = \frac{dx}{dt} = \frac{\frac{dx'}{dt'} \frac{dt'}{dt} + v \frac{dt'}{dt}}{\sqrt{1 - \beta^2}} \quad \text{and} \quad \frac{dt'}{dt} = \frac{1 - \frac{v}{c^2} u_x}{\sqrt{1 - \beta^2}}, \quad (40)$$

that is 
$$u_x = \frac{(u_x' + v) \left(1 - \frac{v}{c^2} u_x\right)}{1 - \beta^2}.$$

Solving for  $u_x$ , 
$$u_x = \frac{u_x' + v}{1 + \frac{v}{c^2} u_x'}. \quad . \quad . \quad . \quad (41)$$

Hence by (40), 
$$\frac{dt'}{dt} = \frac{\sqrt{1 - \beta^2}}{1 + \frac{vu_x'}{c^2}}, \quad . \quad . \quad . \quad (42)$$



and 
$$u_y = \frac{dy}{dt} = \frac{dy'}{dt} = \frac{dy'}{dt'} \frac{dt'}{dt} = \frac{u_y' \sqrt{1-\beta^2}}{1 + \frac{vu_x'}{c^2}} \quad \dots \quad (43)$$

Similarly 
$$u_z = \frac{u_z' \sqrt{1-\beta^2}}{1 + \frac{vu_x'}{c^2}} \quad \dots \quad (44)$$

On account of the denominator of (41), the sum of two velocities which are separately less than  $c$  can never exceed the velocity of light. The Einstein velocity addition theorem gives the simplest explanation of the index of refraction of moving bodies (the Fizeau Effect, p. 477).

Despite the fact that the velocity of light appears to be an upper limit for velocities, it nevertheless cannot be denied that velocities greater than  $c$  are known. In optics, for example, there are media with index of refraction  $n < 1$ . This means that the *wave* velocity  $u$  is greater than  $c$ . But signals can be given only by pulses, or groups of waves, which are propagated with the so-called group velocity (see p. 65). Sommerfeld has shown that this velocity is less than or equal to  $c$  for all media. Another example of such a *geometric* velocity exceeding  $c$  is the following: Let a long ruler make a very small angle  $\theta$  with the  $x$ -axis of a rectangular system of co-ordinates. If, now, the ruler be moved in the  $y$ -direction with a velocity  $v$ , the point of intersection with the  $x$ -axis will move with a speed  $v/\tan \theta$ . It is evident that this velocity can be made to exceed  $c$  by taking  $\theta$  sufficiently small, yet such geometric velocities in excess of  $c$  can never act as connexions between cause and effect.

### (e) Doppler effect and aberration

Let a source of light  $L'$ , at rest in the frame  $\Sigma'$ , send out a spherical wave in this frame. Let the wave be represented by

$$S' = \frac{A'}{r'} e^{2\pi i \nu' (t' - r'/c) + i\delta'}, \quad \dots \quad (45)$$

where all quantities observed in  $\Sigma'$  are denoted by accents. Let this spherical wave meet an observer  $P$  at rest in a system  $\Sigma$ . Let the co-ordinates of  $P$  be

$x$  and  $y$  in  $\Sigma$  and  $x'$  and  $y'$  in  $\Sigma'$ , taking both the  $x$ - and  $x'$ -axis in the direction of the relative motion of  $\Sigma'$  with respect to  $\Sigma$  (fig. 7). In  $\Sigma'$ , let the ray from  $L'$  to  $P$  make an angle  $\theta'$  with the  $x'$ -axis, so that

$$r' = x' \cos \theta' + y' \sin \theta'$$

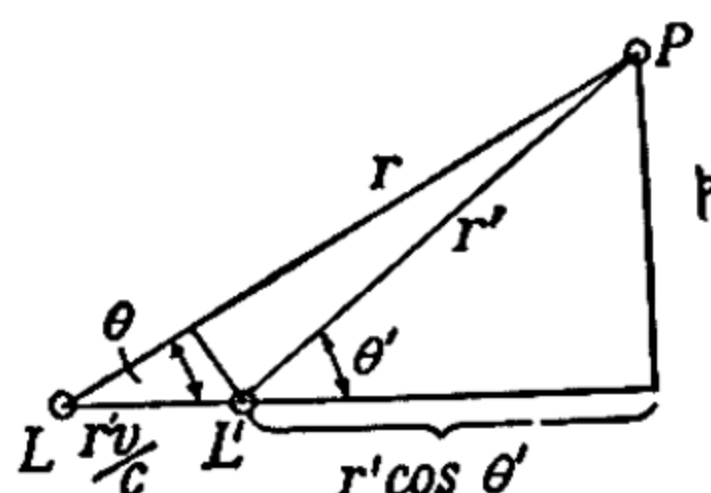


Fig. 7

and hence 
$$S' = \frac{A'}{r'} e^{2\pi i \nu' (t' - (x' \cos \theta' + y' \sin \theta')/c) + i\delta'} \quad \dots \quad (46)$$

The observer in  $\Sigma$  interprets the approaching wave surfaces as coming from a point-source of light belonging to his system, so that the relation which holds for him is

$$S = \frac{A}{r} e^{2\pi i \nu (t - (x \cos \theta + y \sin \theta)/c) + i\delta} \quad \dots \quad (47)$$

Moreover, the co-ordinates and the time in the two systems must be connected by the Lorentz transformation. If we employ these relationships to express the accented quantities in terms of the unaccented ones, then comparison of  $S$  with  $S'$  must yield a relation which is valid for every co-ordinate and for every time. This is possible only if the coefficients of  $x$ ,  $y$  and  $t$  in the exponents are separately equal. Thus from

$$\begin{aligned} 2\pi i \nu' \left( \frac{t - \frac{vx}{c^2}}{\sqrt{1 - \beta^2}} - \frac{(x - vt) \cos \theta'}{c \sqrt{1 - \beta^2}} - \frac{y \sin \theta'}{c} \right) \\ \equiv 2\pi i \nu \left( t - \frac{x \cos \theta}{c} - \frac{y \sin \theta}{c} \right) \quad \dots \quad (48) \end{aligned}$$

we obtain, after reduction,

$$\nu = \nu' \frac{1 + (v/c) \cos \theta'}{\sqrt{1 - \beta^2}}, \quad \dots \quad (49)$$

$$\begin{aligned} \cos \theta = \frac{\cos \theta' + v/c}{1 + (v/c) \cos \theta'}, \quad \sin \theta = \frac{\sin \theta' \sqrt{1 - \beta^2}}{1 + (v/c) \cos \theta'}, \\ \tan \theta = \frac{\sin \theta' \sqrt{1 - \beta^2}}{\cos \theta' + v/c} \quad \dots \quad (50) \end{aligned}$$

Equation (49) gives the exact expression for the optical Doppler effect; this agrees with the acoustic value to terms of second order. It is worth noticing, however, that for transverse observation ( $\theta = \pi/2$ ) there exists a second-order effect here, while in acoustics  $\nu = \nu'$  exactly. By (50),  $\cos \theta' = -v/c$  for  $\theta = \pi/2$ , and hence (49) gives  $\nu = \nu' \sqrt{1 - \beta^2}$ ; thus a shift toward the red takes place. Unfortunately, no experiment thus far attempted has been able even to approach the accuracy needed for the observation of this effect. On the other hand, H. E. Ives was able to increase the accuracy of measurement of the longitudinal effect as far as second-order terms, by observing in both directions along a beam of positive rays and measuring the displacement of the mean with respect to the frequency of the source at rest. By (49), the mean value is

$$\bar{\nu} = \nu' / \sqrt{1 - \beta^2} = \nu' (1 + \frac{1}{2} \beta^2),$$

while from (24) and (25),  $\bar{v} = v'(1 + \beta^2)$ .

The experiment decided in favour of the former value.

The meaning of equations (50) is the following: If we neglect second-order terms, we see from the figure that the source is seen from  $\Sigma$  at the position it occupied when the wave just arriving at  $P$  left the source, i.e. it is seen where it was at a time  $r'/c$  earlier. This is the so-called time of retardation. In this time the source has moved a distance  $vr'/c$ , and from fig. 7 we have approximately

$$\cos \theta = [r' \cos \theta' + r'v/c] / [r' + (r'v/c) \cos \theta].$$

Putting  $\cos \theta' = \cos \theta$  in the small second term of the denominator, we obtain (50). If the two systems remain in uniform rectilinear motion relative to each other we do not notice this displacement, but since the earth reverses the direction of motion in its orbit every six months, the displacement of the apparent position of a star also reverses. This is the phenomenon of *aberration*, first observed by Bradley in 1728.

The ordinary explanation of aberration is based on a ballistic analogy. If the ray of light were a projectile making a hole in the objective and in the eyepiece of a telescope, the line joining the two holes would, on account of the motion of the telescope, deviate from the true direction by the angle computed above. Since aberration is a first-order effect (i.e. one depending on  $v/c$ ), it *must* follow correctly to this approximation from the older mechanics and electrical theory as well.

## 8. Geometric Representation of the Lorentz Transformation. The Four-dimensional World. Calculation with World-vectors.

In order to compare the transformations of Lorentz and Galileo we start with a very simple device—a graphical one-dimensional time-

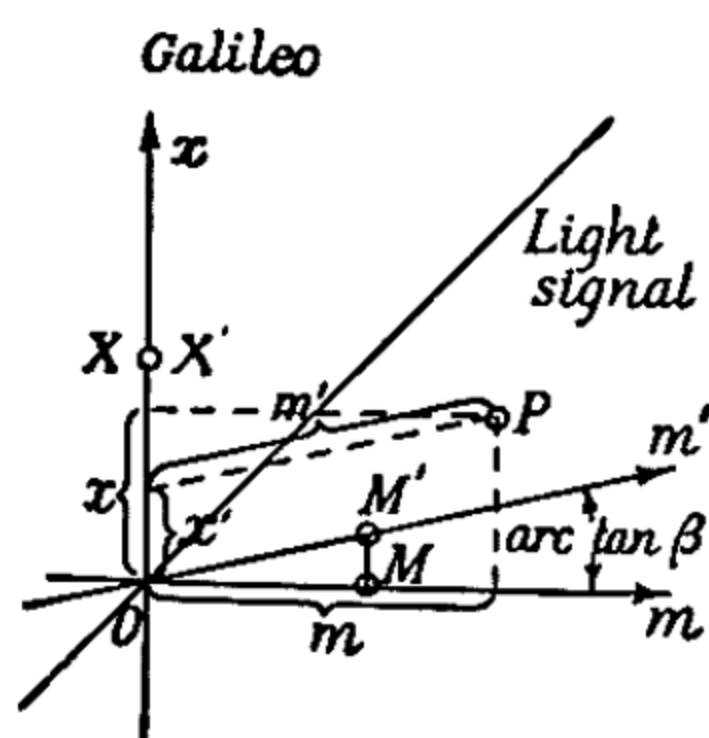


Fig. 8a

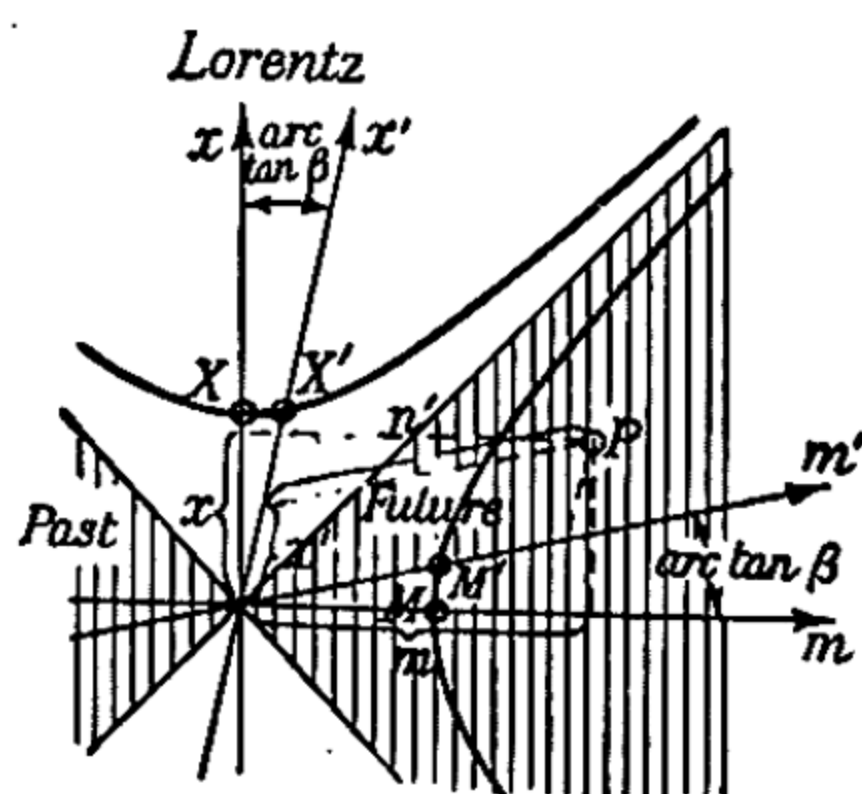


Fig. 8b

table (fig. 8a). We choose as abscissa not the time itself, but the product  $ct$ , which we denote by  $m$ . As ordinate we select the distance  $x$  traversed on a line which we take as the  $x$ -axis. Uniform motion will then be represented in this diagram by a straight line of slope  $\beta = v/c$ .



A light signal is thus represented by a line of slope unity. Now let all quantities be converted to a system moving with velocity  $v$ . If the Galilean transformation  $t' = t$ ,  $x' = x - vt$  holds, this means that we take a new  $m'$ -axis of slope  $\beta$  in the diagram, the  $x$ -axis remaining unchanged. If the unit on the old  $m$ -axis is given by the segment  $OM$ , then the unit on the  $m'$ -axis is given by  $OM'$ , where  $M'$  is the intersection of the  $m'$ -axis with a line through  $M$  parallel to the  $x$ -axis. The new co-ordinate system is oblique, but the  $x$ -axis is the same as before.

The situation is different if the Lorentz transformation holds in place of the Galilean. As we shall show immediately, the new axes and unit segments are obtained as follows: We draw (fig. 8b) the hyperbolas

$$m^2 - x^2 = 1 \quad \text{and} \quad m^2 - x^2 = -1, \quad . \quad . \quad (51)$$

whose asymptotes  $m^2 - x^2 = 0$  correspond to the propagation of light signals. The new  $m'$ -axis is drawn as a line of slope  $\beta$ , as above. The unit point  $M'$ , however, is now the intersection of this line with the hyperbola  $m^2 - x^2 = 1$ . Corresponding to the circumstance that the time must also be transformed in the relativity theory, a new  $x'$ -axis must also be found. This is the line of gradient  $1/\beta$ , i.e. the line symmetric to the  $m'$ -axis with respect to the line of slope unity. The unit along the  $x'$ -axis is determined by the point of intersection  $X'$  of this axis with the hyperbola  $m^2 - x^2 = -1$ . We proceed to show that these constructions make the transformation from the rectangular  $m, x$  system to the oblique  $m', x'$  system equivalent to the Lorentz transformation. First of all, the transformation to oblique co-ordinates with the same origin is given by a linear homogeneous transformation of the co-ordinates:

$$m = \lambda m' + \mu x', \quad x = \nu m' + \rho x'. \quad . \quad . \quad (52)$$

We determine the coefficients by means of the unit points. For  $M'$  we have  $m' = 1$ ,  $x' = 0$ , that is,  $m = \lambda$ . Again, for the abscissa of the intersection of the line  $x = \beta m$  with the hyperbola  $m^2 - x^2 = 1$  we have  $1/\sqrt{1 - \beta^2}$ . Hence  $\lambda = 1/\sqrt{1 - \beta^2}$ . For the  $x$ -co-ordinate of the same point,  $x = \nu = \beta m = \beta/\sqrt{1 - \beta^2}$ . In the same way, from the co-ordinates of the point  $X'$ ,

$$\mu = \frac{\beta}{\sqrt{1 - \beta^2}} \quad \text{and} \quad \rho = \frac{1}{\sqrt{1 - \beta^2}}.$$

$$\text{We thus have} \quad \left. \begin{aligned} m &= \frac{m' + \beta x'}{\sqrt{1 - \beta^2}} & \text{or} & \quad t = \frac{t' + \frac{v}{c^2} x'}{\sqrt{1 - \beta^2}} \\ x &= \frac{x' + \beta m'}{\sqrt{1 - \beta^2}} & \text{or} & \quad x = \frac{x' + vt'}{\sqrt{1 - \beta^2}} \end{aligned} \right\} . \quad . \quad (53)$$

This representation provides an interesting partition of the  $xt$ -continuum. Since the maximum possible velocity for an agency connecting cause and effect is  $c$ , the straight line representing such an instrumentality can have, at most, a slope equal to unity. Hence if in fig. 8b the point  $O$  represents *the present*, then in the shaded area to the right of  $O$ , we find all events which are causally conditioned by  $O$ ; this region therefore represents *the future*. All events which themselves have conditioned  $O$  are found in the shaded area to the left of this point—*the past*. Events whose representative points ("world-points") lie in the unshaded portions of the plane cannot stand in any causal relation with  $O$ . These points, and only these points, may be brought into simultaneity with  $O$  by transforming to a moving system, since we can pass an  $x'$ -axis through such points, thereby making  $m' = 0$ . If we include the other co-ordinates  $y$  and  $z$  of three-dimensional space, the asymptotes become the lateral surface of a four-dimensional cone  $x^2 + y^2 + z^2 - m^2 = 0$ ; this does not change any essential item of our discussion.

There is, however, another way of representing the situation which is much more important for the formal analysis, but which cannot be visualized as well as the above representation, since it employs complex numbers. If we introduce the co-ordinate  $l$  in place of  $m = ct$ , where  $l = ict$ ,

$$\text{we have} \quad x^2 + y^2 + z^2 + l^2 \equiv x'^2 + y'^2 + z'^2 + l'^2. \quad (54)$$

This means that the square of the radius vector remains unchanged under a transformation of co-ordinates in the four-dimensional  $x, y, z, l$  space. As in the three-dimensional case, the only transformation, aside from reflection and translation, which leaves  $r^2$  unaltered is a rotation of the co-ordinate system. Actually, it can be shown that a rotation of the  $x, l$  system of axes yields the Lorentz transformation. If the  $l'$ -axis forms an angle  $\phi$  with the  $l$ -axis, then, according to the familiar relations of analytic geometry,\*

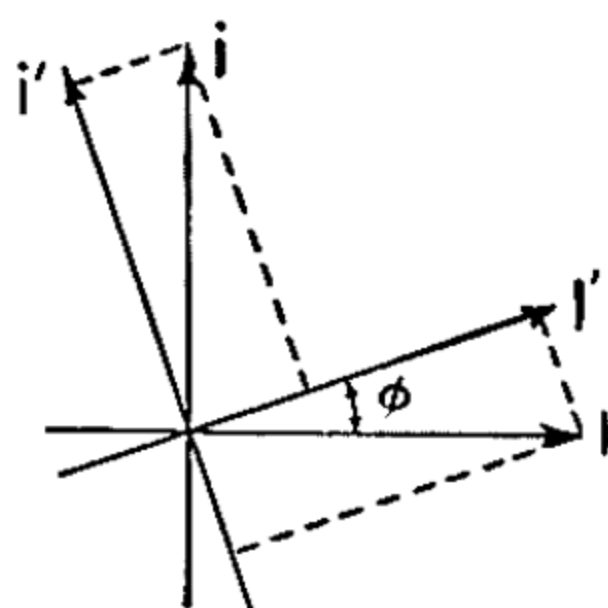


Fig. 9

$$\left. \begin{aligned} l' &= l \cos \phi + x \sin \phi \\ x' &= -l \sin \phi + x \cos \phi \end{aligned} \right\} \quad (55)$$

\* This transformation, corresponding to rotation of a plane two-dimensional system of co-ordinates, may be verified readily by writing the expression for the radius vector in both systems:

$$r = ll + xi = l'l' + x'i'.$$

From fig. 9 we see that

$$l = l' \cos \phi - i' \sin \phi \quad \text{and} \quad i = l' \sin \phi + i' \cos \phi.$$

Inserting these values in the equation for  $r$  and equating the coefficients of  $l'$  and  $i'$  to zero separately, we obtain (55).

If we set

$$\tan \phi = -i\beta, \quad . . . . . (56)$$

that is,

$$\cos \phi = \frac{1}{\sqrt{1 - \beta^2}}, \quad \sin \phi = \frac{-i\beta}{\sqrt{1 - \beta^2}}, \quad . . . (57)$$

we obtain

$$\left. \begin{aligned} x' &= \frac{x + i\beta l}{\sqrt{1 - \beta^2}} = \frac{x - vt}{\sqrt{1 - \beta^2}} \\ l' &= \frac{l - i\beta x}{\sqrt{1 - \beta^2}} \quad \text{or} \quad t' = \frac{t - \frac{v}{c^2}x}{\sqrt{1 - \beta^2}} \end{aligned} \right\}, \quad . . . (58)$$

i.e. the Lorentz transformation.

Thus we have the important relationship discovered by Minkowski, that the transformation from one system of reference to another moving relative to the first with a velocity  $v$  corresponds to a rotation of axes in the four-dimensional  $x, y, z, l$  world, the angle of rotation being given by equation (56). This makes it possible, by extending ordinary vector analysis to four dimensions, to derive generally valid laws in the form of equations between four-dimensional vectors. In what follows, these vectors will be called *world-vectors*, and will be denoted by bars beneath their symbols. Just as the ordinary vector equations representing the laws of Newtonian mechanics may be resolved along any co-ordinate system, so the co-ordinate resolution of world-vector equations yields a representation for a given system of reference. Resolution according to a system turned with respect to the first gives the representation for a reference system in motion relative to the first one.

The following rules hold for calculations with world-vectors: Addition and subtraction are defined as the same operations performed on corresponding components. Graphical addition would, of course, be possible also in four dimensions, according to the methods of descriptive geometry. Since the four-dimensional co-ordinates are also orthogonal,

$$\begin{aligned} \bar{i}\bar{i} = 1, \quad \bar{j}\bar{j} = 1, \quad \bar{k}\bar{k} = 1, \quad \bar{l}\bar{l} = 1, \quad \bar{i}\bar{j} = 0, \quad \bar{i}\bar{k} = 0, \\ \bar{i}\bar{l} = 0, \quad \bar{j}\bar{k} = 0, \quad \&c. \end{aligned} \quad (59)$$

Hence the scalar product of two world-vectors

$$\underline{u} = u_x \bar{i} + u_y \bar{j} + u_z \bar{k} + u_l \bar{l} \quad \text{and} \quad \underline{v} = v_x \bar{i} + v_y \bar{j} + v_z \bar{k} + v_l \bar{l}$$

is defined as

$$\underline{u}\underline{v} = u_x v_x + u_y v_y + u_z v_z + u_l v_l. \quad . . . (60)$$

The definition of the vector product is somewhat more difficult, since this combination really signifies an antisymmetric tensor, and it is more or less fortuitous that such a tensor in three dimensions can be represented by a vector. In four-dimensional space an antisymmetric tensor has six components, a vector has four. We define as the vector



product  $[\underline{u}\underline{v}]$  of two world-vectors the antisymmetric world-tensor  $\Phi$ , whose components are the determinants of the matrix

$$\begin{vmatrix} u_x & u_y & u_z & u_l \\ v_x & v_y & v_z & v_l \end{vmatrix}$$

i.e.

$$\left. \begin{aligned} \Phi_{xy} &= u_x v_y - u_y v_x, & \Phi_{xz} &= u_x v_z - u_z v_x, & \Phi_{xl} &= u_x v_l - u_l v_x, \\ \Phi_{yz} &= u_y v_z - u_z v_y, & \Phi_{yl} &= u_y v_l - u_l v_y, & \Phi_{zl} &= u_z v_l - u_l v_z \end{aligned} \right\}. \quad (61)$$

Corresponding to the three-dimensional  $\Delta$ -operator we have a four-dimensional operator

$$\square = i \frac{\partial}{\partial x} + j \frac{\partial}{\partial y} + k \frac{\partial}{\partial z} + l \frac{\partial}{\partial l}$$

whose combination with world-vectors or scalars gives, in complete analogy with the three-dimensional case, the following differential operations:

$$\underline{\text{grad}} u = \square u = i \frac{\partial u}{\partial x} + j \frac{\partial u}{\partial y} + k \frac{\partial u}{\partial z} + l \frac{\partial u}{\partial l}. \quad (62)$$

$$\text{div } \underline{v} = \square v = \frac{\partial v_x}{\partial x} + \frac{\partial v_y}{\partial y} + \frac{\partial v_z}{\partial z} + \frac{\partial v_l}{\partial l}. \quad (63)$$

$$\text{curl } \underline{v} = [\square v] = \begin{vmatrix} \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} & \frac{\partial}{\partial l} \\ v_x & v_y & v_z & v_l \end{vmatrix}. \quad (64)$$

In place of the equation  $\text{div curl } v = 0$  we have a somewhat more complicated property of world-tensors formed by taking the curl of world-vectors; viz. if  $\Phi = \text{curl } v$ , i.e.

$$\begin{aligned} \Phi_{xy} &= \frac{\partial v_y}{\partial x} - \frac{\partial v_x}{\partial y}, & \Phi_{xz} &= \frac{\partial v_z}{\partial x} - \frac{\partial v_x}{\partial z}, & \Phi_{xl} &= \frac{\partial v_l}{\partial x} - \frac{\partial v_x}{\partial l}, \\ \Phi_{yz} &= \frac{\partial v_z}{\partial y} - \frac{\partial v_y}{\partial z}, & \Phi_{yl} &= \frac{\partial v_l}{\partial y} - \frac{\partial v_y}{\partial l}, & \Phi_{zl} &= \frac{\partial v_l}{\partial z} - \frac{\partial v_z}{\partial l}, \end{aligned}$$

then we have

$$\left. \begin{aligned} \frac{\partial \Phi_{xy}}{\partial z} + \frac{\partial \Phi_{yz}}{\partial x} + \frac{\partial \Phi_{zx}}{\partial y} &= 0 \\ \frac{\partial \Phi_{yz}}{\partial l} + \frac{\partial \Phi_{zl}}{\partial y} + \frac{\partial \Phi_{ly}}{\partial z} &= 0 \\ \frac{\partial \Phi_{zl}}{\partial x} + \frac{\partial \Phi_{lx}}{\partial z} + \frac{\partial \Phi_{xz}}{\partial l} &= 0 \\ \frac{\partial \Phi_{lx}}{\partial y} + \frac{\partial \Phi_{xy}}{\partial l} + \frac{\partial \Phi_{yl}}{\partial x} &= 0 \end{aligned} \right\} \dots \dots \dots (65)$$

### 9. Newton's Second Law in the Theory of Relativity. The Variability of Mass and the Inertia of Energy.

The Galilean transformation forms the basis of Newtonian mechanics. It is our task to modify the laws of mechanics in such a way that they become compatible with the experimentally confirmed Lorentz transformation. It is true that the velocities which are met with in the mechanics of macroscopic bodies are so small that proof of the correctness of the transformation by means of terrestrial experiments has so far not been found possible; but we can impart to electrically charged particles (electrons) velocities approaching  $c$ , which therefore permit of a test. The formulation of the relativistic basis of mechanics is usually approached from electrodynamics, but it is also possible to derive the new laws from purely mechanical considerations. Adopting the latter method of approach, we have to

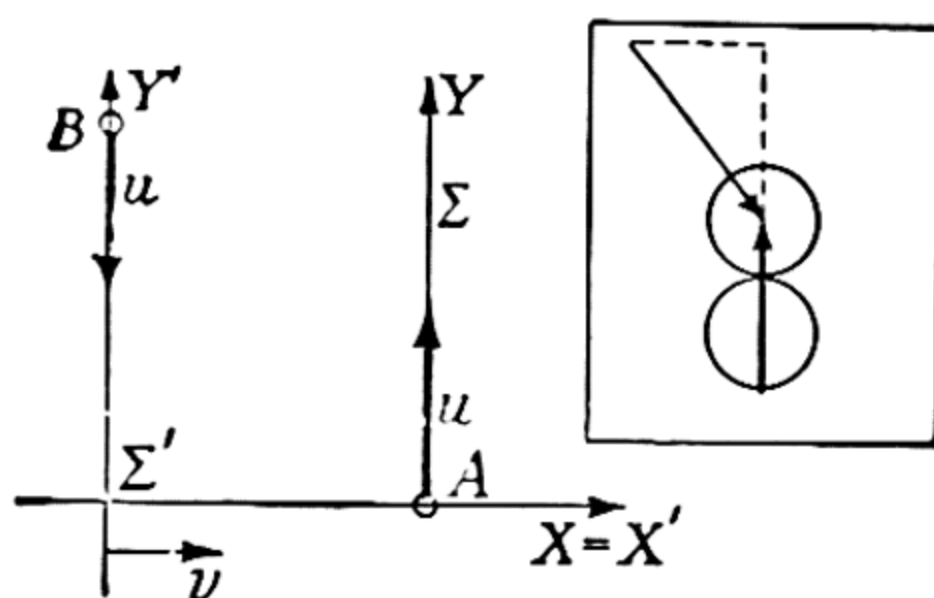


Fig. 10

introduce the concepts of quantity of motion (momentum) and of energy into the mechanics of the Lorentz transformation, and determine the corresponding mathematical expressions in such a way that the laws of conservation of momentum and of energy are retained.

The simplest process whose result is conditioned by these laws of conservation is that of the impact of two equal spheres. Following Tolman, we use this case for a hypothetical experiment. Imagine an observer located on the  $y$ -axis of a frame or system  $\Sigma$  and another observer on the  $y'$ -axis of a second system  $\Sigma'$ . Let the first observer throw a ball along the  $y$ -axis, the second one along the negative  $y'$ -axis (fig. 10). The masses of the two spheres are to be equal, when measured in the same system. Assume the two velocities of projection, as measured each by the corresponding observer, to be equal and opposite. Let a collision take place as a result of the motion of  $\Sigma'$ , and let the instants at which the spheres are thrown be so chosen that the line of centres at the instant of impact has the direction of the  $y$ -axis. Since the spheres are assumed to be smooth, they cannot exert tangential forces on one another, and so the  $x$ -components of the veloci-

ties are unchanged by the impact. If we denote the ball thrown by observer  $A$  in system  $\Sigma$  by  $a$ , the other by  $b$ , then, according to the rule for the addition of velocities, the velocity components before impact are

$$\left. \begin{array}{ll} \text{as seen from } \Sigma: & \text{as seen from } \Sigma': \\ w_{ax} = 0 & w'_{ax} = -v. \\ w_{ay} = u & w'_{ay} = u \sqrt{1 - \frac{v^2}{c^2}} \\ w_{bx} = v & w'_{bx} = 0 \\ w_{by} = -u \sqrt{1 - \frac{v^2}{c^2}} & w'_{by} = -u \end{array} \right\}. \quad (66)$$

Let all quantities *after* impact be denoted by bars over the symbols. In order to be able to apply the law of the conservation of momentum we make the assumption that, in relativistic mechanics also, the momentum vector has the direction of the velocity vector. Here, however, we replace the scalar factor which represents the constant mass of the sphere in Newtonian mechanics by a variable multiplier  $\alpha$ . Being a scalar, this can depend only upon a scalar quantity, say the magnitude  $w$  of the velocity. Under the above hypotheses, the  $x$ -component in the system  $\Sigma$  yields

$$\begin{aligned} \alpha(\bar{u}) \cdot 0 + \alpha \left( \sqrt{v^2 + \bar{u}^2 \left(1 - \frac{v^2}{c^2}\right)} \right) \cdot v \\ = \alpha(u) \cdot 0 + \alpha \left( \sqrt{v^2 + u^2 \left(1 - \frac{v^2}{c^2}\right)} \right) \cdot v, \end{aligned} \quad (67)$$

or

$$\bar{u} = \pm u. \quad \dots \dots \dots (68)$$

Newtonian mechanics requires that  $\bar{u} = -u$  in this case. Since relativistic mechanics must reduce to the Newtonian form for small velocities, we must use the negative sign here. Further, the  $y$ -component yields the equation

$$\begin{aligned} \alpha(\bar{u}) \cdot \bar{u} - \alpha \left( \sqrt{v^2 + \bar{u}^2 \left(1 - \frac{v^2}{c^2}\right)} \right) \cdot \bar{u} \sqrt{1 - \frac{v^2}{c^2}} \\ = \alpha(u) \cdot u - \alpha \left( \sqrt{v^2 + u^2 \left(1 - \frac{v^2}{c^2}\right)} \right) \cdot u \sqrt{1 - \frac{v^2}{c^2}}. \end{aligned} \quad (69)$$

This is satisfied for  $\bar{u} = -u$  if

$$\alpha \left( \sqrt{v^2 + u^2 \left(1 - \frac{v^2}{c^2}\right)} \right) = \frac{\alpha(u)}{\sqrt{1 - \frac{v^2}{c^2}}}. \quad \dots \dots (70)$$



If we now let the velocity of projection become smaller and smaller,

$$\alpha(v) = \frac{\alpha(0)}{\sqrt{1 - \frac{v^2}{c^2}}} \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad (71)$$

For very small velocities, however, we must arrive at the Newtonian expressions, in which the scalar factor  $\alpha$  is the mass  $m_0$ . If we call this scalar factor the *mass* in the relativity theory also, then the result states that the mass increases with increasing velocity. Denoting the mass corresponding to a state of rest by  $m_0$ , we obtain the fundamental relation

$$m = \frac{m_0}{\sqrt{1 - \frac{v^2}{c^2}}} \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad (72)$$

In what follows, the quantity  $m$  will be more accurately termed the *momental mass* or the *apparent mass*.

We now formulate the fundamental law of mechanics in the original Newtonian form, in which the force is set equal to the time rate of change of momentum:

$$\mathbf{F} = \frac{d}{dt}(m\mathbf{v}) = \frac{d}{dt} \frac{m_0 \mathbf{v}}{\sqrt{1 - \frac{v^2}{c^2}}} \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad (73)$$

This differs from the old form in that  $m$  may no longer be placed before the sign of differentiation, since it is now variable.

According to this equation, the *inertial mass*, i.e. the resistance offered by a body to acceleration, is not now identical with the apparent mass; in fact the inertial mass in a frame of reference in which the body has a speed  $v$  will actually have different values in the two cases when the acceleration takes place in the direction of the initial velocity and normal thereto. To show this, we take the  $x$ -axis of our co-ordinate system in the direction of the velocity, while the acceleration is taken to be in the  $xy$ -plane, i.e.

$$\frac{dx}{dt} = v_x = v, \quad \frac{dy}{dt} = v_y = 0, \quad \frac{dz}{dt} = 0; \quad \frac{d^2x}{dt^2} = \frac{dv}{dt} = a_x, \quad \frac{d^2y}{dt^2} = a_y, \quad \frac{d^2z}{dt^2} = 0.$$

Carrying out the differentiation in equation (73) and putting  $v_x = v$ , we have

$$F_x = \frac{d}{dt} \frac{m_0 v_x}{\sqrt{1 - \frac{v^2}{c^2}}} = \frac{m_0}{\sqrt{1 - \frac{v^2}{c^2}}} a_x + \frac{m_0 v_x \frac{v}{c^2}}{\left(\sqrt{1 - \frac{v^2}{c^2}}\right)^3} a_x = \frac{m_0}{\left(\sqrt{1 - \frac{v^2}{c^2}}\right)^3} a_x, \quad (74)$$

$$F_v = \frac{d}{dt} \frac{m_0 v_v}{\sqrt{1 - \frac{v^2}{c^2}}} = \frac{m_0}{\sqrt{1 - \frac{v^2}{c^2}}} a_v. \quad \dots \quad (75)$$

Thus the inertial mass occurring as the factor of proportionality in the equation connecting force and acceleration has the magnitude

$$\frac{m_0}{\left(\sqrt{1 - \frac{v^2}{c^2}}\right)^3}$$

for the  $x$ -component, and the magnitude

$$\frac{m_0}{\sqrt{1 - \frac{v^2}{c^2}}}$$

for the  $y$ -component. This means that the force vector and the acceleration vector have different directions; the acceleration makes a greater angle with the velocity than the force does, since the longitudinal inertial mass is greater than the transverse inertial mass.

The formulation of the Second Law given by equation (73) presupposes a particular frame of reference. For a frame in motion relative to this frame, the quantities entering in (73) must be transformed according to the Lorentz transformation. This is unsatisfactory from the standpoint of the relativity theory. A generally valid relationship ought to have the form of an equation connecting world-vectors, which may be resolved (i.e. split up into components) in any four-dimensional co-ordinate system, so that the relationship remains valid for any given frame, or for any frame in motion relative to it. We now attempt to formulate such relationships. First, we note that at any given instant one special frame is marked out from all others, viz. that which has the same instantaneous velocity  $\mathbf{v}$  as the particle. We shall call this the rest-frame. The time in this frame we call the *local time* or *proper time*  $\tau$ . Hence for a frame in which the velocity of the particle is  $v$ , and the time is  $t$ , we have, by equation (36) (p. 243),

$$\frac{dt}{d\tau} = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}}. \quad \dots \quad (76)$$

Take now the four-dimensional radius vector by which the position and time of the particle are specified:

$$\underline{r} = xi + yj + zk + ll,$$

where, as at (54), p. 249,  $l = ict$ . By differentiating with respect to

the proper time, we obtain a new world-vector  $\underline{q}$  which we call the *four-velocity*. For a particle at rest, this vector has the value  $\underline{q} = icl$ , on account of the vanishing of the derivatives  $dx/d\tau$ ,  $dy/d\tau$  and  $dz/d\tau$ . The square of the magnitude of the vector  $\underline{q}$  which is independent of the co-ordinate system, is  $-c^2$ , a constant for all motions. In another four-dimensional system of co-ordinates with time  $t$ , in which the particle has the velocity  $v$ ,

$$\underline{q} = \frac{d\mathbf{r}}{dt} \frac{dt}{d\tau} = \left\{ \frac{dx}{dt} \mathbf{i} + \frac{dy}{dt} \mathbf{j} + \frac{dz}{dt} \mathbf{k} + icl \right\} \cdot \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}} \quad (77)$$

The spatial components of the world-vector are thus given in this system by the components of the ordinary vector

$$\frac{\mathbf{v}}{\sqrt{1 - \frac{v^2}{c^2}}}$$

and the time component ( $l$ -component) is given by

$$\frac{ic}{\sqrt{1 - \frac{v^2}{c^2}}}$$

Hence in the above form of the Second Law, in (73), p. 254, we can put the spatial components of  $\underline{q}$ , combined into a vector  $\mathbf{q}_r$ , in place of  $\mathbf{v} / \sqrt{1 - \frac{v^2}{c^2}}$ :

$$\mathbf{F} = \frac{d}{dt} (m_0 \mathbf{q}_r) = \sqrt{1 - \frac{v^2}{c^2}} \frac{d}{d\tau} (m_0 \mathbf{q}_r). \quad \dots \quad (78)$$

We now introduce in place of the force  $\mathbf{F}$  a world-vector  $\underline{\mathbf{P}}$ —the “four-force”—whose spatial components are identical with those of  $\mathbf{F}$  in the rest-frame, but which satisfy the relationship

$$\underline{\mathbf{P}}_r = \frac{\mathbf{F}}{\sqrt{1 - \frac{v^2}{c^2}}} \quad \dots \quad (79)$$

in another reference frame. Then the Second Law has the familiar form

$$\underline{\mathbf{P}} = \frac{d}{d\tau} (m_0 \underline{\mathbf{q}}). \quad \dots \quad (80)$$

This form leads to a further important result.



If, in analogy with the method of proving the theorem of the conservation of energy in classical mechanics, we take the scalar product of both members of (80) with  $\underline{q}$ , we obtain

$$\underline{P} \underline{q} = \frac{d}{d\tau} (m_0 \underline{q}) \underline{q} = \frac{1}{2} \frac{d}{d\tau} (m_0 \underline{q}^2) = 0. \quad . \quad . \quad . \quad (81)$$

The right side vanishes, since the magnitude of  $\underline{q}$  is the constant  $ic$ . If the left member be written in components, the spatial components being combined into ordinary vectors, we have

$$P_r q_r + P_i q_i = 0,$$

i.e. 
$$\left( \frac{\mathbf{F}}{\sqrt{1 - \frac{v^2}{c^2}}} \right) \left( \frac{\mathbf{v}}{\sqrt{1 - \frac{v^2}{c^2}}} \right) + P_i \frac{ic}{\sqrt{1 - \frac{v^2}{c^2}}} = 0.$$

The product  $\mathbf{F} \mathbf{v}$  is the work done per unit time. If we replace  $P_i$  on the left side by the value

$$P_i = \frac{d}{d\tau} \left( \frac{m_0 ic}{\sqrt{1 - \frac{v^2}{c^2}}} \right) = \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}} \frac{d}{dt} \left( \frac{m_0 ic}{\sqrt{1 - \frac{v^2}{c^2}}} \right)$$

obtained from equation (80), we obtain

$$\mathbf{F} \dot{\mathbf{v}} = \frac{d}{dt} \frac{m_0 c^2}{\sqrt{1 - \frac{v^2}{c^2}}} = \frac{dN}{dt}, \text{ say.}$$

The quantity  $N$  introduced here has the value

$$N = \frac{m_0 c^2}{\sqrt{1 - \frac{v^2}{c^2}}} = m_0 c^2 + \frac{1}{2} m_0 v^2 + \frac{3}{8} m_0 \frac{v^4}{c^2} + \dots \quad (82)$$

to which a constant of integration must still be added. If we give this constant the value  $-m_0 c^2$ , we see that the development of the radical is a first approximation to the kinetic energy  $T$  of Newtonian mechanics. Thus the conservation of energy retains its usual form in relativity: the work done per second is equal to the increase in kinetic energy per unit time.

What is the meaning of the quantity  $m_0 c^2$ , which has the dimensions of energy? If we take the constant of integration to be zero, the term  $m_0 c^2$  represents the energy of the particle at rest. This energy has an enormous value. It is natural to interpret this to mean that inertial mass represents a large accumulation of energy and that,

reciprocally, to every amount of energy  $E$  there corresponds an inertial mass

$$m_0 = \frac{E}{c^2} \quad . \quad . \quad . \quad . \quad . \quad . \quad (83)$$

This is the *Principle of the Inertia of Energy*.\*

According to this law, the inertial mass of a body is increased whenever energy is supplied (e.g. by heating). The energy available in the laboratory, however, is far too small to make this increase of mass observable in any way, but it is otherwise with the enormous energy changes involved in nuclear processes (see Chap. XLII). The loss in mass of the stars, resulting from radiation of energy, is also considerable (cf. *Ex.* 67).

*Ex.* 67. What is the annual loss in mass of the sun, if approximately 2 calories of radiated energy are received by each square centimetre of the earth's surface per minute? (The distance from the earth to the sun is about  $150 \times 10^6$  km.)

## 10. Fundamental Concepts of the Generalized Theory of Relativity.

The Lorentz transformation refers only to frames of reference which are in uniform rectilinear motion with respect to each other. We have seen that the assumption of an absolute space is superfluous and in contradiction with experience for this type of motion. But the situation is very much more difficult for frames in accelerated motion, e.g. in rotation. The phenomena accompanying rotation do seem to point to the existence of an absolute space, for which alone Newton's Second Law is valid; for the occurrence of inertial forces, such as centrifugal or Coriolis forces, indicates that the bodies are in accelerated motion with respect to such a space. But is it not more advisable to say that the accelerated motion takes place with respect to the remaining masses of the universe? If these masses, and not a cosmic ether, are the cause of centrifugal forces, then such forces would have to vanish if the remaining material of the cosmos were removed. But, unfortunately, this is an experiment which one cannot perform. Nevertheless, as early as 1896, B. and J. Friedländer undertook experiments to detect the existence of feeble centrifugal forces on stationary bodies placed near large, rapidly rotating masses, the latter being intended to take the place, in a certain measure, of the masses of the

\* The history of this far-reaching law is, briefly, as follows: As early as 1887, Voigt took up the discussion of the time transformation in a little-noticed paper, and in 1892 Lorentz revived the question in connexion with the optics of moving media. In 1900 Larmor wrote the Lorentz transformation in the form we know it to-day, and four years later Lorentz published a paper in which he treated the optics of moving media entirely from the standpoint of the Lorentz transformation. In 1905 Poincaré and Einstein independently arrived at substantially the conclusions given in § 7. Einstein's paper contained the mass-energy relation. Shortly before this, Hasenöhl, in Austria, treated the special case of a radiation-filled enclosure, showing that the contained radiation would oppose any acceleration of the container as though the radiation had a virtual mass given by (83).



universe. On account of the disparity of the masses in question, the experiment would be doomed to failure even now. Even so, we may claim that a theory which explains inertial forces through the action of matter has the advantage of not introducing a hypothetical space (cosmic ether), whose existence has not been verified in the realm of uniform motion. But the path which leads to a strict theory based upon these ideas is long and arduous. In the first place, we must give up something which we are naturally inclined to take for granted: the validity of Euclidean geometry in our four-dimensional space-time continuum. This geometry, which was fundamental in the special theory of relativity, appeared in the expression for the square of the radius vector  $r^2 = x^2 + y^2 + z^2 + l^2$  (Theorem of Pythagoras).

The idea that the physical *space* in which we live might not have the properties of Euclidean geometry is not new. Riemann expounded the view that the metric of space is a physical property, and not a matter of intuition. As long as we measure small triangles on the surface of the earth with ordinary rough apparatus, we find that all the propositions of Euclidean geometry—in particular that concerning the sum of the angles of a triangle—are valid. However, if we deal with larger areas, the straight lines are replaced by chains stretched over the earth's surface, and more exact measurement reveals that the sum of the angles of a triangle is greater than  $\pi$ . In exactly the same way, a triangle in space, whose sides are formed, say, by light rays, might have an angle sum different from  $\pi$ ; in fact, Gauss tried in this way, though without success, to find such a departure from the Euclidean metric.

From the epistemological standpoint, the idea here advanced means nothing less than a reintroduction of the ether, since we are really once more ascribing physical properties to space. But we can dispense with this physical space by using the formula: "It is mere assumption to regard space-time as Euclidean. By suitable hypotheses with regard to the metric of the space-time frame, a more perspicuous representation of the world may be obtained."

Einstein's fundamental idea in the generalized theory of relativity was to determine the metric of the space-time continuum by means of the distribution and velocity of matter in such a way that in a frame of reference relative to which the matter is in accelerated motion the inertial forces result automatically. In this connexion, use is made of the equivalence of gravitational fields and accelerations mentioned on p. 231. As explained there, this equivalence is based on the constant ratio ( $=1$ ), of inertial mass to gravitational mass for all bodies. This equality is by no means a matter of course. Nevertheless, it forms the basis of the generalized theory of relativity, just as the negative result of the experiment to detect an ether-wind forms the basis of the special theory. In the generalized relativity theory the



fundamental law of mechanics is: A particle (or light ray) upon which no electromagnetic forces act describes a line of stationary length (geodesic, or generalized straight line) in the space-time continuum, the metric of the continuum being determined by the distribution and velocities of the material bodies. This connexion is mathematically very complicated and cannot be expounded in detail here; the expression for a geodesic in the non-Euclidean metric is also complicated. The new feature is that gravitation, which is not included in electromagnetic effects, appears in this manner as a property of space, and that the same centrifugal forces would arise whether the earth is considered to turn on its axis or the rest of the universe to be rotating about this axis. Hence the conflict between the points of view of Ptolemy and Copernicus vanishes.

There are three cases where the generalized theory of relativity leads to small deviations from the results of the old physics. The calculations yield the result that the perihelia of the planetary orbits advance by very small amounts. For the planet Mercury, where the effect is greatest, the amount of this advance is computed to be only 43'' per century. Again, on account of the change in the metric in the neighbourhood of heavy bodies, a light ray passing near the sun on its way to us from a distant star will suffer a small deviation, so that the position of a star whose line of sight grazes the sun will be somewhat displaced. This effect can be looked for at the time of a total solar eclipse. Finally, the change in the time metric caused by gravitation causes spectral lines originating in a strong gravitational field (e.g. that of the sun) to be shifted very slightly toward the red, when compared by an outside observer with lines generated in his own system. All three effects exist. However, it is a matter of extreme difficulty to allow correctly for all disturbing influences, so that we are not yet justified in speaking of a *quantitative* verification of the generalized theory of relativity.

Since the idea of determining the metric of the space-time continuum in such a way that the laws of nature are automatically contained therein has been successful in the case of gravitation, it was natural to attempt to extend the notion to electromagnetic phenomena also. Einstein, by means of a still more general geometry, actually was able to devise a unified field theory which includes electromagnetism. Nevertheless, it must be recognized that such a correlation embraces only one aspect of the universe; it fails to explain atomic structure—the electrons, protons and quanta.

With the large particle speeds now available in the laboratory, the special theory of relativity has become indispensable in atomic and nuclear research. However, up to the present time at least, the general theory has been applied only to cosmological physics, and more specific connexions with experiment are still lacking.

## PART III

### FIELD THEORY OF ELECTROMAGNETIC AND OPTICAL PHENOMENA

Two different methods are employed in the theoretical treatment of electrical and magnetic phenomena. The first method starts with macroscopically measurable quantities and describes mathematically the relationships between such quantities found by experiment. An enormous range of phenomena may be brought within the scope of this method in the form of a system of differential equations. The integration of these equations for particular cases then permits us to answer, by rigorous calculation, a large number of questions, including many of technical importance. This method employs only experimental laws and their mathematical consequences; it is therefore entirely free of hypothesis, and cannot come into conflict with experience. Since it makes no assumptions whatsoever concerning the structure of electrical charges or of matter, we call this the continuum theory, or the field theory. Nevertheless, it was soon recognized that a large number of very striking phenomena, e.g. those of electrolysis, are entirely left out when this point of view is taken. In order to explain these phenomena, we must make assumptions concerning the structure of matter and of electrical charges. We cannot verify these assumptions by direct observation, but only by testing their consequences. This second method will be called the atomistic method. While the continuum theory will be treated first, we shall nevertheless not attempt to exclude all reference to atomistic concepts; on the contrary, we shall occasionally avail ourselves of a side-glance at the atomistic picture of the phenomena for greater vividness.





## CHAPTER XI

## THE ELECTROSTATIC FIELD IN A VACUUM (OR IN AIR)

## 1. Definitions.

(a) *Electrical field strength and charge*

In order to arrive at a definition of electrical field strength we make use of an arrangement whose real purpose we shall not learn until later (p. 425). Two horizontal metal plates (fig. 1) are placed opposite each other at a distance of a few centimetres and are held in this position by amber supports. By means of an atomizer, a cloud of very fine droplets of oil is blown into the space between the plates. The drops are illuminated from the side and may be observed in a telescope.




Fig. 1

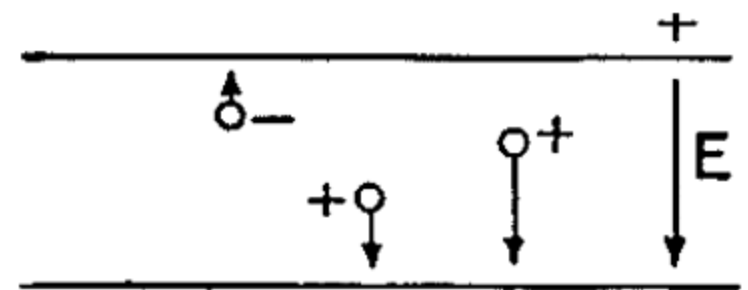


Fig. 1

We observe that, under the combined effect of gravity and air resistance (cf. Stokes's theory, p. 217), the drops fall with uniform velocity. We now touch the upper plate with a glass rod which has been rubbed with silk, the lower plate being connected to the earth by a wire. It is observed that the motion of the droplets changes: some rise, some fall with various speeds, and others remain suspended in space. If we concentrate attention on a single drop, we observe that it changes its velocity now and then, especially if some radium is placed near by. The direction of motion, however, is always normal to the plates. We conclude from this observation that a new force, whose cause is connected with the touching of the upper plate by the glass rod, acts on the droplets. The space between the plates has thus become the seat of a *field of force* which is superimposed upon the forces already present. We find further that the magnitude of the force and its algebraic sign depend on the state of the droplet. Hence we set the force vector  $\mathbf{F}$  equal to the product of a vector  $\mathbf{E}$ , independent of the state of the drop, and a number  $e$  which we call the charge \* on the drop:

$$\mathbf{F} = e\mathbf{E}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

$E$  is called the electrical field strength. By observing the changes in

\* The droplet becomes charged when it is formed by atomization of the oil.

the operating force, as revealed by changes in the velocity of rise or fall of the drop, we conclude that the droplets alter their charges under the influence of the radium radiations.

The algebraic sign of the charge is conventionally determined in such a way that a particle repelled by the upper plate—the one touched by the glass rod—is said to carry a positive charge. In the light of our present knowledge, this convention is not a happy one, since the most important charged particle, the electron, is thus given the negative sign.

### (b) *Lines of force and flux*

Suppose that we have determined the field strength  $\mathbf{E}$  at every point of space by means of a test body whose charge we arbitrarily take to be unity. We can then specify the field strength at every point by a vector. However, such a representation does not afford a very good general view of the whole field. It is better to draw curves whose tangents give the direction of the field at any point. Then the strength of the field must be specified in some other way. We still have a choice as to how close together we shall draw these curves, the so-called *lines of force*. We adopt a method which gives the density of the lines of force as a measure of the field strength. The lines of force of the electrostatic field form an orthotomic system, i.e. they are the orthogonal trajectories of a family of surfaces. This is not true of all vector fields. At every point of space there is a certain plane, which is tangential to the orthogonal surface at that point. We arrange matters so that the number of lines of force passing through each unit area of the orthogonal surface is proportional to the field strength. The number of lines of force passing through an arbitrary area is called the *flux* or *induction*  $\Phi$  through that area. As may be seen from fig. 16 (p. 23), the flux through an arbitrarily oriented element of area  $d\mathbf{S}$  amounts to

$$d\Phi = \mathbf{E}d\mathbf{S}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

Hence the flux through any finite surface  $S$  is

$$\Phi = \int_S \mathbf{E}d\mathbf{S}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

## 2. Electrical Charge (Quantity of Electricity) as the Source of Flux.

If we take the integral  $\oint \mathbf{E}d\mathbf{S}$  over a closed surface  $S$ , this gives the excess of the number of lines of force leaving the volume bounded by  $S$  over the number entering. If, in the experimental arrangement of § 1, we take the surface  $S$  so that it does not intersect the plates, we find that the integral vanishes: \* that is, no lines of force originate

\* With the proviso, of course, that there are now no charged particles between the plates.

in the space between the plates. The situation is different if our surface encloses, say, the upper plate. The surface integral now has a definite value, and we can connect the non-vanishing of this integral with the fact that the plate was touched by the rubbed glass rod. We naturally conjecture that we have thus given the plate an electric charge, and can verify this directly by bringing the plate into a similar system of larger dimensions; we shall then recognize the presence of a charge by the resulting forces, just as in the case of the oil drops. We take the value of the integral as a measure of the charge within the enclosed region, and set

$$e \propto \oint \mathbf{E} d\mathbf{S}, \quad . . . . . (4)$$

the units being still undetermined. The electrical charge or quantity of electricity thus plays a double rôle: an active part as the generator of a field, and a passive part as the object upon which the field acts. The *density of charge* at a point  $P$  of space is the limit of the ratio of the charge within a volume element containing  $P$  to the volume of this element:

$$\rho = \lim \frac{\Delta e}{\Delta \tau} \propto \frac{\oint \mathbf{E} d\mathbf{S}}{\Delta \tau} = \text{div } \mathbf{E}. \quad . . . . . (5)$$

That is, at every point of space the divergence of the electrical field strength is proportional to the density of charge at that point. The limit of  $\Delta e/\Delta \tau$  is not always finite; in nature we find surface distributions of charge as well as volume distributions. If we construct a cylinder whose base is  $\Delta S$ , and whose generators are normal to a charged surface, we can make the altitude of the cylinder small of a higher order than the base without changing the amount of electricity enclosed by it. In this instance we define the surface density of charge  $\sigma$  to be the limit of  $\Delta e/\Delta S$ , i.e. the amount of electricity on unit *area* of the surface. Let us take the surface integral of the electrical field strength over such a small cylinder (cf. p. 43, fig. 27). We distinguish the two sides of the surface by the subscripts 1 and 2, and suppose the normal to be directed from 1 towards 2. Then, since the lateral surfaces are small of a higher order than  $\Delta S$ , we have

$$\Delta S(\mathbf{n} \cdot \mathbf{E}_2 - \mathbf{n} \cdot \mathbf{E}_1) \propto \Delta e. \quad . . . . . (6)$$

If the limit of  $\Delta e/\Delta S$  is not zero,  $\mathbf{n} \cdot \mathbf{E}_2 - \mathbf{n} \cdot \mathbf{E}_1$  must be different from zero. This means that the normal component of the field strength,  $\mathbf{n} \cdot \mathbf{E}$ , has different values on the two sides of a charged surface, i.e. there is a discontinuity in the field strength. The expression  $\mathbf{n} \cdot (\mathbf{E}_2 - \mathbf{E}_1)$  is called the surface divergence of  $\mathbf{E}$  (written  $\text{Div } \mathbf{E}$ ), as on p. 44. We may thus say that *for a surface distribution of*



charge, the surface density is proportional to the surface divergence of  $\mathbf{E}$ .

$$\text{Div } \mathbf{E} \propto \sigma. \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

To bring the quantities  $e$ ,  $\mathbf{E}$  into connexion with the absolute system of units, we start with the field of a single point charge  $e$ . From symmetry we see that the lines of force are radial and of equal density in all directions. The orthogonal surfaces in this case are spheres with centre at  $e$ . The induction through a sphere of radius  $r$  amounts to

$$\Phi = \oint_{\text{Sphere}} \mathbf{E} d\mathbf{S}.$$

Since the density of the lines of force is the same everywhere,  $|\mathbf{E}|$  is constant over the surface of the sphere and

$$\Phi = \oint \mathbf{E} d\mathbf{S} = 4\pi r^2 |\mathbf{E}| = \gamma e, \quad \text{i.e. } |\mathbf{E}| = \frac{\gamma e}{4\pi r^2},$$

where  $\gamma$  is a factor of proportionality. The force on a second charge  $e'$  becomes

$$|\mathbf{F}| = \frac{\gamma e e'}{4\pi r^2}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

This is *Coulomb's Law*.\*

The units are determined by specifying the factor of proportionality  $\gamma$ . If we place  $\gamma = 4\pi$ , we obtain  $e$  in the absolute electrostatic system. In addition, however, there are three other systems of units in use. To avoid confusion it is desirable to use only two systems. One is the C.G.S., introduced above, in which all electrical quantities are measured in electrostatic units (e.s.u.) and all magnetic quantities in electromagnetic units (e.m.u.), to be introduced later. The other is the rationalized *metre-kilogram-second* (M.K.S.) system, sometimes called the Giorgi system after one of its first proponents. The fundamental mechanical and electrical units in this system are the metre, kilogram, second, ampere and volt, the size of the electrical units being given by their legal definitions in terms of the amount of silver deposited electrolytically and the electromotive force of a standard cell, respectively. The M.K.S. unit of force is called one *newton*, and is equal to  $10^5$  dynes. The C.G.S. system is still indispensable in atomic physics, for all numerical constants are specified in terms of these units; but

\* If we examine the familiar diagram giving the lines of force in the neighbourhood of two equal and opposite point charges, we see that all the lines of force originating at the positive charge converge into the negative. Hence the field is no longer spherically symmetrical, as it was before we introduced the second charge, yet we used this property of the field in calculating the force between two charges. The contradiction disappears if, in calculating the force on a given charge, we draw only the lines of force emanating from (or converging on) the other charge. The field pattern mentioned above gives the forces acting upon a third charge introduced into the field.

the M.K.S. system has such great advantages in macroscopic problems that its use is becoming more and more widespread. In using M.K.S. units, however, it must be remembered that even quantities such as the number of molecules per mol differ from the familiar C.G.S. values, for in the M.K.S. system the kilogram mol is used in place of the gram mol.

For comparison, the more important electrical formulas will be given here in both systems. The M.K.S. unit of quantity of electricity is the ampere-second (coulomb), and that of field strength is the volt/metre. Thus the relation between charge and field strength as given by (4) contains a dimensional factor of proportionality, called the *permittivity* (or *capacitivity*)\* of free space:

$$K_0 = 8.859 \times 10^{-12} \text{ amp. sec./metre volt.}$$

According to (8), we then have, in

C.G.S. units ( $\gamma = 4\pi$ )		M.K.S. units ( $\gamma = 1/K_0$ )
$\oint \mathbf{E} d\mathbf{S} = 4\pi e. \quad . \quad . \quad . \quad (4a)$		$e = K_0 \oint \mathbf{E} d\mathbf{S}. \quad . \quad . \quad . \quad (4b)$
$\text{div } \mathbf{E} = 4\pi\rho. \quad . \quad . \quad . \quad (5a)$		$\rho = K_0 \text{div } \mathbf{E}. \quad . \quad . \quad . \quad (5b)$
$\text{Div } \mathbf{E} = 4\pi\sigma. \quad . \quad . \quad . \quad (7a)$		$\sigma = K_0 \text{Div } \mathbf{E}. \quad . \quad . \quad . \quad (7b)$
$F = \frac{ee'}{r^2}. \quad . \quad . \quad . \quad (8a)$		$F = \frac{ee'}{4\pi K_0 r^2}. \quad . \quad . \quad . \quad (8b)$

### 3. The Electrostatic Potential.

The work done by the field when the charge  $e$  experiences a displacement  $d\mathbf{S}$  is

$$dW = \mathbf{F} d\mathbf{S} = e\mathbf{E} d\mathbf{S}. \quad . \quad . \quad . \quad . \quad (9)$$

If  $e$  is moved a finite distance along a curve from  $P_0$  to  $P_1$ , the total work done is

$$W = \int_{P_0}^{P_1} e\mathbf{E} d\mathbf{S} = e \int_{P_0}^{P_1} \mathbf{E} d\mathbf{S}. \quad . \quad . \quad . \quad . \quad (10)$$

*This work is found to be independent of the path  $C$  over which the charge moves from  $P_0$  to  $P_1$ .* On account of this fact, the integral  $\oint \mathbf{E} d\mathbf{S}$  taken over a closed path is zero, i.e.

$$\oint \mathbf{E} d\mathbf{S} = 0. \quad . \quad . \quad . \quad . \quad (10')$$

This means that *the field strength  $\mathbf{E}$  is irrotational*. Now we know from vector analysis that every irrotational vector may be repre-

\*The terms *dielectric constant* and *specific inductive capacity* refer to material dielectrics.

sented as the gradient of a scalar. This means a great simplification: instead of giving the numerical values of three components at each point, it is sufficient to give the numerical value of one scalar at every point. Then by taking the gradient, i.e. by pure differential operations, the vector components may be found. The field considered above is of this kind, for in this field

$$dW = \text{grad } W ds,$$

and by equation (9), for  $e = 1$ ,

$$dW = \mathbf{E} ds, \quad \text{i.e. } \mathbf{E} = \text{grad } W.$$

However, it is more convenient to work with the quantity  $(-W)$ , in analogy with potential energy in mechanics. This would be the work done *against* the field when unit charge moves. This quantity is called the electrostatic potential  $V$ .\* It is to be noted that the *product*  $eV$ , not  $V$  itself, has the dimensions of work. In terms of this concept we have

$$\mathbf{E} = -\text{grad } V; \quad \int_{P_0}^{P_1} \mathbf{E} ds = V_0 - V_1. \quad \dots \quad (11)$$

Taking the divergence of the field strength, we have by (5a)

$$\text{div grad } V = \Delta V = -4\pi\rho, \quad \text{or} \quad K_0 \Delta V = -\rho. \quad \dots \quad (12)$$

This differential equation for the electrostatic potential is known as *Poisson's Equation*. In particular, at all points of a region devoid of charge,

$$\Delta V = 0, \quad \dots \dots \dots (13)$$

which is Laplace's Equation.

On account of the importance of the equations of Poisson and Laplace, we give the expressions for  $\Delta V$  in the commonest co-ordinate systems (cf. *Ex. 22*, p. 43):

In rectangular co-ordinates:

$$\Delta V = \frac{\partial^2 V}{\partial x^2} + \frac{\partial^2 V}{\partial y^2} + \frac{\partial^2 V}{\partial z^2}. \quad \dots \dots \dots (14)$$

In cylindrical co-ordinates:

$$\Delta V = \frac{\partial^2 V}{\partial \rho^2} + \frac{1}{\rho} \frac{\partial V}{\partial \rho} + \frac{1}{\rho^2} \frac{\partial^2 V}{\partial \phi^2} + \frac{\partial^2 V}{\partial z^2}. \quad \dots \dots (15)$$

In spherical co-ordinates:

$$\Delta V = \frac{\partial^2 V}{\partial r^2} + \frac{2}{r} \frac{\partial V}{\partial r} + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2 V}{\partial \phi^2} + \frac{1}{r^2} \frac{\partial^2 V}{\partial \theta^2} + \frac{1}{r^2} \cot \theta \frac{\partial V}{\partial \theta}. \quad (16)$$

\* More accurately, it is the difference in potential between the end points  $P_0$  and  $P_1$  of the path. Cf. equation (11).



*Ex. 68.* Is it possible to produce a uni-directional electrostatic field in which the magnitude of the field strength increases in a direction normal to  $\mathbf{E}$ ?

*Ex. 69.* At the surface of the earth the electrostatic field in the atmosphere is 100 volts/m., and at a height of 1500 m. the field strength is 25 volts/m., the field being directed toward the centre of the earth. What is the surface charge of the earth and what is the mean space charge in the atmosphere between 0 and 1500 m. altitude?

*Ex. 70.* In a gas discharge between parallel plates, the potential is measured as a function of the distance from the cathode, using probes. Explain how the field strength and the distribution of space charge can be found from these results.

#### 4. Simple Examples of the Electrostatic Field in a Vacuum (or in Air).

The problem of calculating an electrostatic field may be considered solved when the distribution of the potential has been found, for the field is determined from the potential by taking the gradient, i.e. by simple differentiation. The problem is to find solutions of the differential equation for the potential which are in agreement with the particular conditions of the problem in hand. This problem may be solved at once if the distribution of charges in the entire space is given. But this is generally not the case. We must take into consideration the fact that there are substances—so-called conductors of electricity—within which no electrostatic field can be maintained, but upon which the charges may move freely. Every surface of a conductor must therefore be a level surface of the field, and we must find a solution of  $\Delta V = 0$  which satisfies the boundary conditions imposed by the positions of the conductors. It is only after this has been done that we can determine the distribution of charge by evaluating  $\text{Div grad } V$ . Under these conditions the problem is considerably more difficult to solve.

##### A. No Conductors in the Field.

###### (a) Potential due to point charges

We begin with the simplest instance of a given distribution of charge, viz. the simple point charge. Its field is spherically symmetrical, and has already been calculated in § 2 (p. 266). The corresponding potential might be set down at once, but we wish to proceed in a somewhat more general way. By equation (16) (p. 268), the equation for the potential in a space devoid of charge (Laplace's Equation) is, for spherical symmetry,

$$\Delta V = \frac{d^2 V}{dr^2} + \frac{2}{r} \frac{dV}{dr} = 0. \quad \dots \dots (17)$$

The integral is

$$V = \frac{A}{r} + B. \quad \dots \dots (18)$$

$B$  is the value of the potential at infinity. We normalize the potential in such a way that it vanishes at infinity. This can always be done, since any system of charges in the finite region appears to be concentrated at a single point when viewed from infinity. Hence we may specify, as a boundary condition of the differential equation, that  $V$

vanishes at least to the order  $1/r$  at infinity. The constant  $A$  is determined by taking the surface integral over a sphere with centre at the point charge. The value thus found by equation (4a) (p. 267) is  $e$ , while (4b) yields  $e/4\pi K_0$ .

If several point charges are given, their fields are superimposed without interference, and so the combined potential is the sum of the potentials due to the separate charges:

$$V = \sum \frac{e_i}{r_i}, \text{ or (in M.K.S. units) } \sum \frac{e_i}{4\pi K_0 r_i}. \quad (18a)$$

For a continuous distribution of charge in space or upon surfaces the sum becomes

$$V = \int \frac{\rho d\tau}{r} + \int \frac{\sigma dS}{r}, \text{ or } V = \int \frac{\rho d\tau}{4\pi K_0 r} + \int \frac{\sigma dS}{4\pi K_0 r}. \quad (19)$$

This equation expressing the independent superposition of the fields may be derived directly from the differential equation with the help of Green's theorem. We give the procedure for the part due to a space distribution of charge. Let  $V$  and  $\psi$  be two scalar point functions which are continuous, together with their first and second spatial derivatives, in a given region. By applying Gauss's theorem to the vectors  $V \text{ grad } \psi$  and  $\psi \text{ grad } V$  in turn, and subtracting, we have\*

$$\oint (V \text{ grad } \psi) d\mathbf{S} - \oint (\psi \text{ grad } V) d\mathbf{S} = \int V \Delta \psi d\tau - \int \psi \Delta V d\tau. \quad (20)$$

This is Green's theorem.

Let  $V$  satisfy the differential equation  $\Delta V = -4\pi\rho$ . Let us try the effect of taking the function  $1/r$  as the value of  $\psi$ , where  $r$  is the distance of the element of charge  $\rho d\tau$  from the point  $P_0$  at which we wish to evaluate  $V$  (cf. fig. 2). Then this function will be discontinuous at  $P_0$ , and we must remove  $P_0$  from the region of integration. This is accomplished by drawing a small sphere of radius  $a$  with  $P_0$  as centre. The surface of this sphere is a part of the bounding surface of the region of integration. According to the convention adopted in the

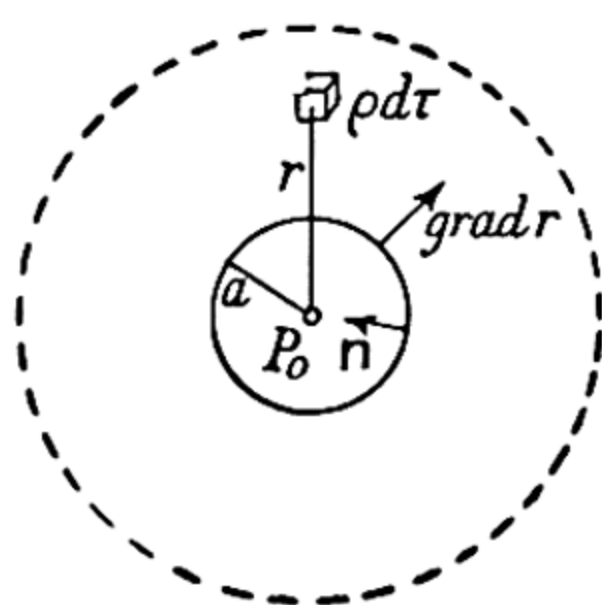


Fig. 2

\* Gauss's theorem (p. 24) is  $\oint \mathbf{v} d\mathbf{S} = \int \text{div } \mathbf{v} d\tau$ .

Here  $\mathbf{v} = V \text{ grad } \psi$ . By equation (88) (p. 38),

$$\begin{aligned} \text{div } (V \text{ grad } \psi) &= (\text{grad } \psi) (\text{grad } V) + V \text{ div grad } \psi \\ &= (\text{grad } \psi) (\text{grad } V) + V \Delta \psi. \end{aligned}$$

Similarly, for the function  $\psi \text{ grad } V$ ,

$$\text{div } (\psi \text{ grad } V) = (\text{grad } \psi) (\text{grad } V) + \psi \Delta V.$$

Substituting in Gauss's theorem and subtracting, we obtain equation (20).





In seeking a solution for the interior, which again can depend only upon  $r$ , it must be remembered that the additive constant in the potential has already been determined by the solution for the exterior; the two solutions must fit at  $r = a$ , since a discontinuity in  $V$  would mean infinitely great field strength. The field strength in the interior is again determined by the flux of force, by equation (4a), p. 267, to be

$$\mathbf{E}_I = \frac{4\pi r^3 \rho}{3r^2} \mathbf{r}_0. \quad . \quad . \quad . \quad . \quad . \quad (24)$$

The potential is obtained with the proper constant if we start with the equation

$$V_I - V_a = \int_{r=a}^{r=0} \mathbf{E} d\mathbf{s} = \int_{r=a}^{r=0} \frac{4\pi r \rho}{3} dr.$$

By (23'), this gives

$$V_I = \frac{4\pi a^2 \rho}{3} + \frac{2\pi a^2 \rho}{3} - \frac{2\pi r^2 \rho}{3} = 2\pi \rho \left( a^2 - \frac{r^2}{3} \right). \quad . \quad (25)$$

It may be noticed that the part of  $V_I$  which depends on  $r$  has the same form, except for sign, as the potential of the force occurring in simple harmonic motion (*Ex.* 35, p. 88). If we assume a negative charge at  $P$ , then the algebraic sign is also the same, and this charge will execute harmonic vibrations in the interior of the sphere.

### (c) Potential due to a dipole

A dipole or doublet is a system consisting of a negative charge  $-e$  and an equal positive charge  $+e$  separated by the elementary segment  $d\mathbf{s}$ . The product  $e d\mathbf{s}$  is assumed to have a finite value  $m$  called the moment or strength of the dipole, i.e. if the distance apart  $d\mathbf{s}$  is made smaller, the charge  $e$  is to increase in such way that the product  $e d\mathbf{s}$  remains finite. As follows from its definition, the dipole moment is a vector. If we denote the distance of any point  $P$  from the negative charge by  $r_-$  and that from the positive charge by  $r_+$  we have

$$V = e \left( -\frac{1}{r_-} + \frac{1}{r_+} \right). \quad . \quad . \quad . \quad . \quad (26)$$

The change in the point function  $1/r$  is obtained by taking the gradient.

It must be noted, however, that in taking the gradient the "field point"  $P$  is moved, while the equipotential surfaces remain fixed. Thus we are not permitted to move the charge. Writing simply  $r$  for  $r_-$  in the difference, we

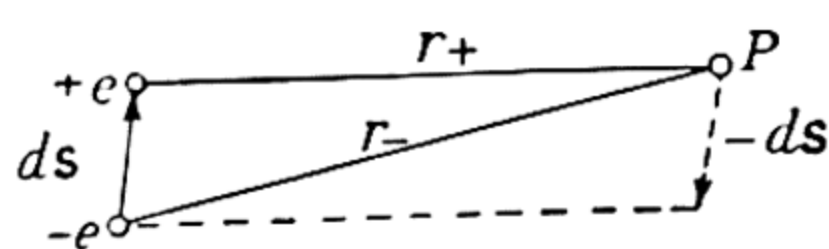


Fig. 3

obtain the following expression by moving the field point (cf. fig. 3):

$$\frac{1}{r_+} = \frac{1}{r_-} - ds \left( \text{grad} \frac{1}{r} \right) = \frac{1}{r_-} + \frac{ds \boldsymbol{r}}{r^3}, \quad \cdot \cdot \cdot (27)$$

whence

$$V = \frac{eds \boldsymbol{r}}{r^3} = \frac{m \boldsymbol{r}}{r^3} = \frac{|m|}{r^2} \cos(m \boldsymbol{r}), \text{ or } V = \frac{m \boldsymbol{r}}{4\pi K_0 r^3}. \quad (28)$$

Thus the potential of a dipole decreases as  $1/r^2$  and depends on the angle which the radius vector makes with the axis of the doublet. The field is therefore no longer spherically symmetrical, but is rotationally symmetrical about the dipole axis. As in all the formulæ occurring here,  $\boldsymbol{r}$  is directed toward the field point.

#### (d) Potential of a double layer

A double layer is a surface covered with dipoles, each having its axis in the direction of the normal to the surface. All dipoles are assumed to be oriented so that one kind of pole is on one side of the surface. Let the dipole moment per square centimetre be  $m$ . The potential at a field point  $P$  is, by (28),

$$V = \int \frac{|m| \cos(n \boldsymbol{r})}{r^2} dS. \quad \cdot \cdot \cdot \cdot (29)$$

If the layer is of uniform density ( $|m|$  constant),

$$V = |m| \int \frac{\cos(n \boldsymbol{r}) dS}{r^2}. \quad \cdot \cdot \cdot \cdot (29')$$

The solid angle  $d\Omega$  subtended by a surface element  $dS$  as seen from  $P$  is defined as the area cut out on a unit sphere drawn about  $P$  by the cone determined by  $dS$ , and this has exactly the same magnitude as the integrand of (29'). Hence

$$V = |m| \Omega. \quad \cdot \cdot \cdot \cdot (29'')$$

According to the convention regarding the direction of  $\boldsymbol{r}$ ,  $V$  is positive if  $P$  faces the positive side of the surface. Up to now we have considered only discontinuities in field strength, but for a double layer the potential itself suffers a discontinuity and so the field would become infinite in the layer if the latter really were infinitely thin. This, of course, never happens in practice.

Consider next a closed double layer, say in the form of a sphere. If elementary cones are drawn from an external point, each one must cut the layer twice, and the cosine of the angle between  $\boldsymbol{n}$  and  $\boldsymbol{r}$  will have opposite sign at the two places ( $\boldsymbol{n}$ , as usual, is drawn outward at all points). Thus the contributions nullify each other in pairs, so that outside the surface the potential is zero. For an internal point,

each nappe of an elementary cone cuts the surface only once; on integrating, the whole surface of the unit sphere is covered. The sign of the contribution is negative, however, since  $\mathbf{n}$  and  $\mathbf{r}$  are opposite in direction. Then

$$V_E - V_I = 4\pi |m|, \quad . . . . . (30)$$

or using the surface gradient, this may be written

$$\text{Grad } V = 4\pi m, \text{ or } K_0 \text{ Grad } V = m. \quad . . (30')$$

This equation holds also for unclosed surfaces. Any unclosed surface may be considered changed to a closed one by supplying an additional portion. The potential of the added part suffers no discontinuity in passing through the original surface, and so contributes nothing to (30'). Double layers arise, for example, when a metallic plate is dipped into a salt solution. The term "electromotive force" as used to denote the corresponding discontinuity of potential is obsolescent.

### *B. Conductors present in the Field.*

#### *(a) Spherical condenser*

Let there be given two concentric conducting spheres of radii  $a$  and  $b$  ( $b > a$ ). Assume a charge  $e$  to be placed upon the inner sphere. The outer sphere is to be held at a potential  $V_b$ , say by means of a conductor connecting it to earth. We may call this potential zero. Naturally, the spherically symmetrical solution of the potential equation is again valid here, but the constants have values different from the previous ones. We have

$$V = \frac{A}{r} + B.$$

By means of the total flux, the value of  $A$  is again found to be  $e$ ; the value of  $B$  is found by imposing the condition that  $V = V_b$  for  $r = b$ , whence

$$V = \frac{e}{r} - \frac{e}{b} + V_b, \text{ or } V = \frac{e}{4\pi K_0 r} - \frac{e}{4\pi K_0 b} + V_b. \quad (31)$$

Hence for the inner sphere,

$$V_a = V_b + e \left( \frac{1}{a} - \frac{1}{b} \right), \text{ or } V_a = V_b + \frac{e}{4\pi K_0} \left( \frac{1}{a} - \frac{1}{b} \right). \quad (32)$$

Now the electrical *capacity* of a system is defined as the ratio of the charge to the potential difference of the bounding conductors. Hence, for the spherical condenser,

$$C = \frac{e}{V_a - V_b} = \frac{ab}{b - a}, \text{ or (M.K.S.) } C = \frac{4\pi K_0 ab}{b - a}. \quad (33)$$



If the radius of the outer sphere is allowed to become infinite we obtain

$$C = a, \text{ or } C = 4\pi K_0 a. \quad . . . . . (34)$$

That is, the capacity of a single isolated sphere, expressed in electrostatic units, is equal to its radius in centimetres. In the M.K.S. system, the capacity of a sphere of 1 cm. radius turns out to be  $1.11 \times 10^{-12}$  farad, the farad being the practical unit of capacity.

(b) *Point charge near an infinite plane conductor at zero potential. Electrical images*

Let the conductor be the  $xy$ -plane (fig. 4), and let the point charge be at  $z = a$ . In the space above the plate, the potential satisfies the condition that  $V$  vanishes on the  $xy$ -plane, and that for every surface enclosing the point charge,

$$\oint (\text{grad } V) d\mathbf{S} = -4\pi e.$$

In the space below the plate, it is required that  $V$  vanish at the  $xy$ -plane, and also that the integral  $\oint (\text{grad } V) d\mathbf{S}$  vanish for every closed surface, since there is no charge in this region. These requirements

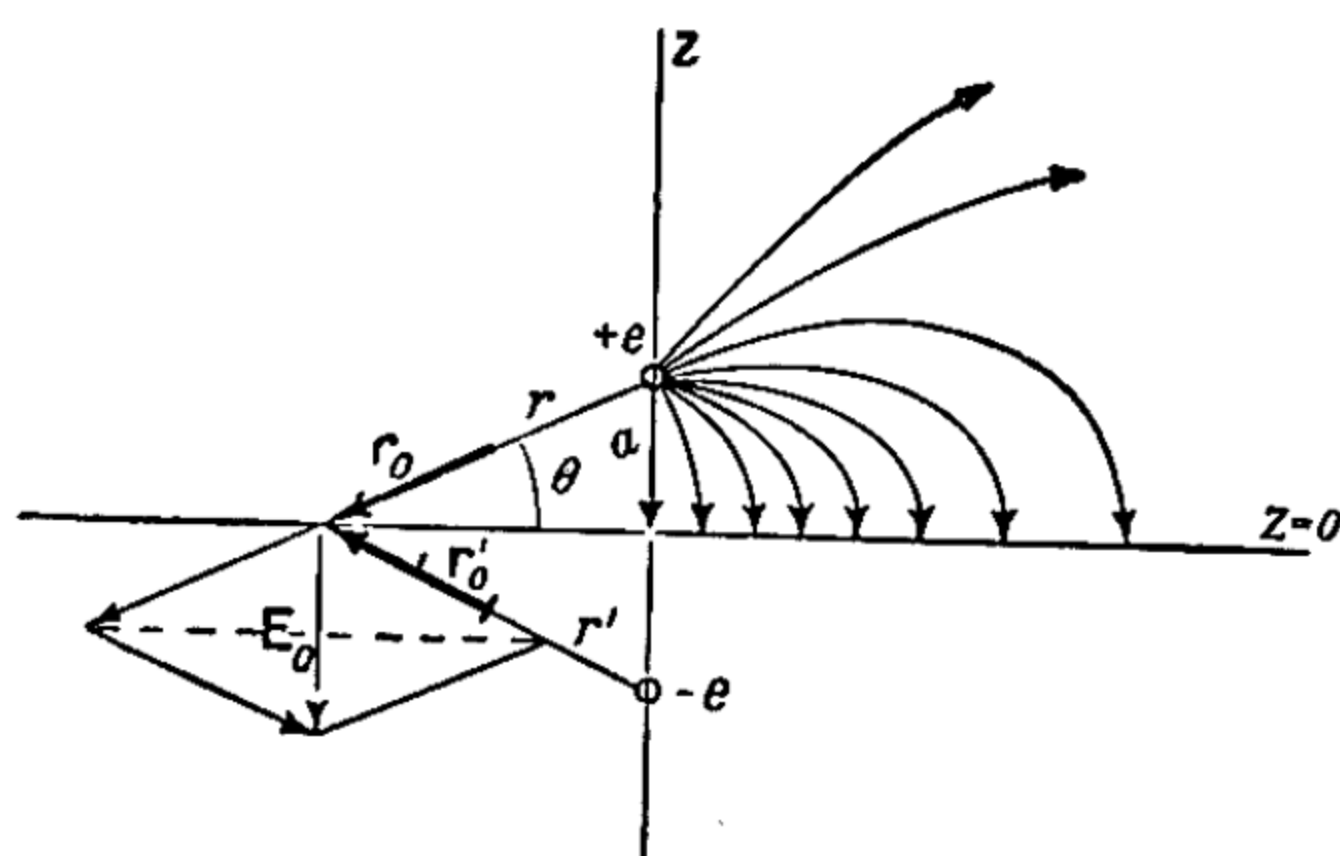


Fig. 4

are fulfilled if we take  $V = 0$  at all points below the plane. In the upper space  $V$  must also vanish at the  $xy$ -plane, but  $\oint \mathbf{E} d\mathbf{S}$  must be equal to  $4\pi e$  for every closed surface round the charge. In this case the solution may be given readily; viz. if we place a charge  $-e$  at the "image-point" of the original charge, the field due to these two charges satisfies all the required conditions in the upper space. Hence in the upper region

$$V = e \left( \frac{1}{r} - \frac{1}{r'} \right). \quad . . . . . (35)$$

This expression, which is a sum of potentials due to point charges, is certainly a solution of the differential equation for the potential. Moreover, for the  $xy$ -plane,  $r = r'$ , and so  $V$  vanishes there. Finally, closed surfaces drawn in the upper space must either contain  $e$  or else contain no charge whatsoever, since the charge  $-e$  lies below the conducting plane. All lines of force originating on  $e$  terminate on the  $xy$ -plane. Since there is no field below this plane, the field strength is discontinuous at the  $xy$ -plane, and so a charge must be present on this plane. In order to determine this charge we first calculate the field strength at the surface of the conducting plane:

$$\mathbf{E} = -\text{grad } V = \frac{e}{r^2} \mathbf{r}_0 - \frac{e}{r'^2} \mathbf{r}'_0. \quad . \quad . \quad . \quad (36)$$

Since  $r = r'$  at the surface of separation, the resultant is a vector  $\mathbf{E}_0$ , normal to the  $xy$ -plane, having the negative  $z$ -direction and of magnitude (see fig. 4)

$$|\mathbf{E}_0| = \frac{2e}{r^2} \sin \theta = \frac{2e}{r^3} a. \quad . \quad . \quad . \quad . \quad (36')$$

Hence by equation (7a) (p. 267), the surface density of charge is

$$\sigma = -\frac{ea}{2\pi r^3}. \quad . \quad . \quad . \quad . \quad . \quad (37)$$

By computing the integral  $\oint \sigma dS$  we may readily verify that the total charge induced on the conducting plane is equal to  $-e$ , corresponding to the fact that the entire flux emanating from  $+e$  terminates on the conductor.

*Ex. 71.* Starting from first principles, compute the force of attraction between the two plates of a parallel-plate condenser, each carrying a charge of magnitude  $e$ . (Make use of the lines-of-force picture and neglect disturbances due to the edges of the conductors.)

*Ex. 72.* Following the method used for a spherical condenser, calculate the capacity of a cylindrical condenser consisting of two coaxial cylinders of length  $l$  and radii  $a$  and  $b$  ( $b > a$ ). Find also the capacity of a parallel-plate condenser consisting of two equal plates of area  $S$  placed at a distance  $d$  apart. Neglect edge effects.

*Ex. 73.* A point charge  $e$  external to a conducting sphere is at a distance  $R$  from the centre. The sphere is held at zero potential. Find the potential at any point in the resulting field. (As in the case of the plane, the effect of the sphere may be replaced by that of a virtual charge within it.)

*Ex. 74.* What is the nature of the equipotential surfaces at a singular point in the field where  $\mathbf{E} = 0$ ? Take the origin of a co-ordinate system at this point and develop the potential in a power series in the co-ordinates. Consider especially the case of rotational symmetry.

## CHAPTER XII

### THE ELECTROSTATIC FIELD IN DIELECTRIC MEDIA

#### 1. Formal Introduction of the Concepts "Dielectric Displacement" and "Free Charge". Boundary Conditions at the Surface of Separation of Two Dielectrics.

We start with the following experimental arrangement. A parallel plate condenser is connected to a quadrant electrometer which indicates the potential difference between the plates. First, let the substance between the plates be air. If we now introduce into the space between the conductors a slab of insulating material, such as a flat piece of hard rubber or a glass plate, the potential difference will be observed to decrease: upon removing the slab, the potential will return to its former value. According to our definition of capacity, this quantity has been increased by introducing the insulator. Since the introduction of the insulating material could not have altered the charge on the conductors, we are led to conclude that the connexion between field strength and charge, found in Chap. XI, § 2 (p. 265), must be modified if the space in which the field operates is filled with a medium other than air. Indeed, we find even a small effect due to the air if we evacuate the space between the plates, but this effect is relatively so small that it may generally be neglected. However, we shall take the results of Chap. XI to be strictly valid only for a vacuum.\*

The ratio of the field strength in vacuum to that in an insulator (or dielectric), for the same distribution of charge, is called the specific inductive capacity or *dielectric constant* of the medium, i.e.

$$\mathbf{E}_v = K\mathbf{E}. \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad (1)$$

We now introduce a new vector  $\mathbf{D}$ , the *dielectric displacement* vector. The same relationships are to hold for this vector as for the vector  $\mathbf{E}_v$  in a vacuum, viz.

$$\left. \begin{array}{l} \oint \mathbf{D} d\mathbf{S} = 4\pi e, \quad \text{div } \mathbf{D} = 4\pi\rho, \quad \text{Div } \mathbf{D} = 4\pi\sigma, \\ \text{while in a vacuum, } \oint \mathbf{E}_v d\mathbf{S} = 4\pi e, \quad \text{div } \mathbf{E}_v = 4\pi\rho, \quad \text{Div } \mathbf{E}_v = 4\pi\sigma. \end{array} \right\} \quad (2)$$

In the M.K.S. system, the factor  $1/K_0$  appears in place of  $4\pi$  on the right side of the equations.

\* The effects for a vacuum and for air differ by less than 0.06 of 1 per cent at standard conditions.



On account of (1), the connexion between  $\mathbf{D}$  and  $\mathbf{E}$  is given by

$$\mathbf{D} = K\mathbf{E}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

This simple relation between  $\mathbf{D}$  and  $\mathbf{E}$  is valid only for isotropic substances, i.e. amorphous materials and regular crystals. For certain anisotropic dielectrics, generally of crystalline structure, the relation (1) must be modified. In such substances the field strength  $\mathbf{E}$  is a linear function of the vacuum field strength, i.e. the dielectric constant is no longer a scalar, but a tensor. We shall, however, restrict our considerations at present to isotropic substances.

On the other hand, the field within a dielectric (which actually may be measured, say by the force on a small test body) may be considered to be the field of certain fictitious charges which would produce the same field in a vacuum. Such charges are called "free" or "apparent" charges. These charges, which we shall denote by accents, are given by the equations

$$\oint \mathbf{E} d\mathbf{S} = 4\pi e', \quad \text{div } \mathbf{E} = 4\pi \rho', \quad \text{Div } \mathbf{E} = 4\pi \sigma'. \quad . \quad (4)$$

True charges occur either as space charges or as charges on the surface of conductors. On the other hand, free charges are found also on the surface of separation of two dielectrics, as we shall see at once. Since no true charges exist at the interface of two dielectrics, we have

$$\text{Div } D = K_1 \mathbf{E}_n^{(1)} - K_2 \mathbf{E}_n^{(2)} = 0. \quad . \quad . \quad . \quad . \quad (5)$$

But this means that  $\text{Div } \mathbf{E}$  is different from zero, since it is only the products of the normal components of  $\mathbf{E}$  by the respective  $K$ 's which are equal. Hence *the surface of separation of two dielectrics is the seat of apparent field-producing charges*. To secure information regarding the tangential components of  $\mathbf{E}$  we make use of the empirical fact that the electrostatic field is everywhere irrotational, even if there are several contiguous dielectrics in the region. If, then, we carry a unit charge over a closed path drawn near the surface of separation of two dielectrics, and with one branch on each side of this surface (fig. 1), the line integral must vanish. This means, however, that the tangential components of  $\mathbf{E}$ , which alone contribute to the integral, must be equal on the two sides:

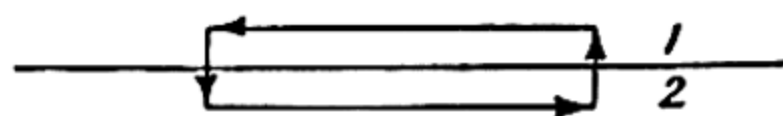


Fig. 1

$$\mathbf{E}_{\text{tan}}^{(1)} = \mathbf{E}_{\text{tan}}^{(2)}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

This equation may be written also in the form

$$[n(\mathbf{E}_2 - \mathbf{E}_1)] = \text{Curl } \mathbf{E} = 0. \quad . \quad . \quad . \quad . \quad (6')$$

The continuity of the tangential component of the electric field strength at the surface of separation of two dielectric media is not limited to

electrostatic fields, but is a general property of the electric field, and will be justified later (p. 344) in a much more general way.

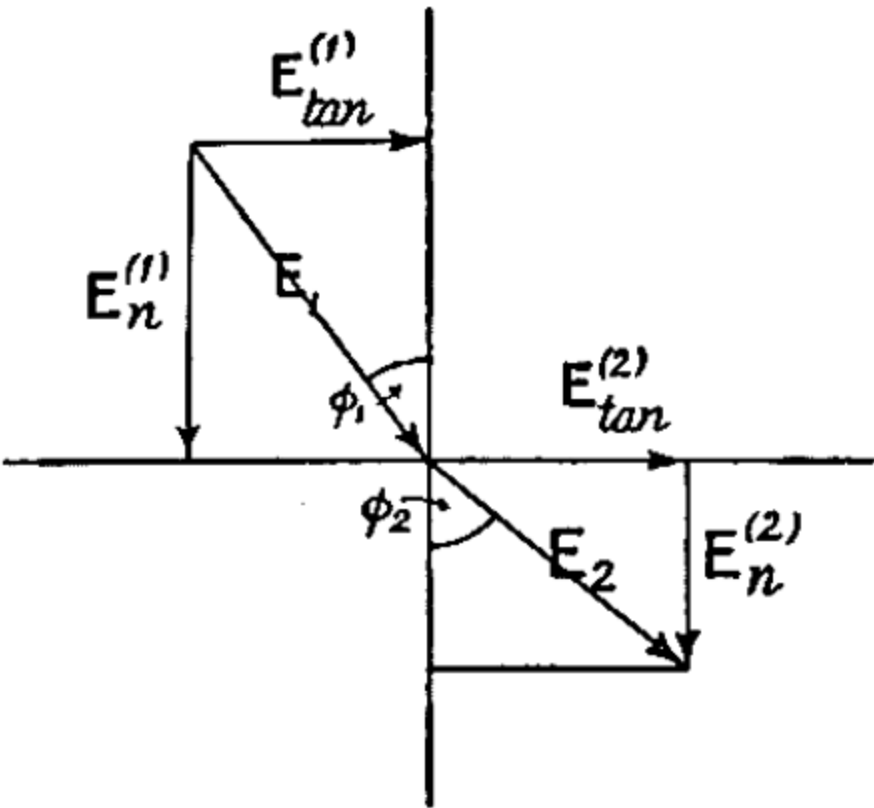


Fig. 2

The two boundary conditions (5) and (6) furnish us with the law of refraction of lines of force. If we consider a plane (fig. 2) determined by the tangent to a line of force in the first medium and the normal to the interface, it follows from (6) that  $E_2$  must also lie in this plane. Further, if  $\phi$  is the angle between the tangent to the line of force and the normal to the interface, it follows from  $\text{Div}(KE) = 0$  that

$$K_1 E_1 \cos \phi_1 = K_2 E_2 \cos \phi_2,$$

and from  $\text{Curl } E = 0$  that

$$E_1 \sin \phi_1 = E_2 \sin \phi_2.$$

Dividing one equation by the other gives the law of refraction for lines of electric force:

$$\frac{\tan \phi_1}{\tan \phi_2} = \frac{K_1}{K_2} \dots \dots \dots (7)$$

2. Polarization of Dielectrics.

We begin with a homogeneous dielectric medium in which a uniform field  $E$  exists. The density of the field lines of the vector  $D$  is  $K$  times that of the field lines of  $E$ . We now take a small cylinder whose generators are parallel to  $E$  and remove the contained material of dielectric constant  $K$ , at the same time placing charges on the bases of the cylinder in such a way that the field is unaltered. Inside the cylinder we now have  $D = E$ , while outside this region we still have  $D = KE$ .

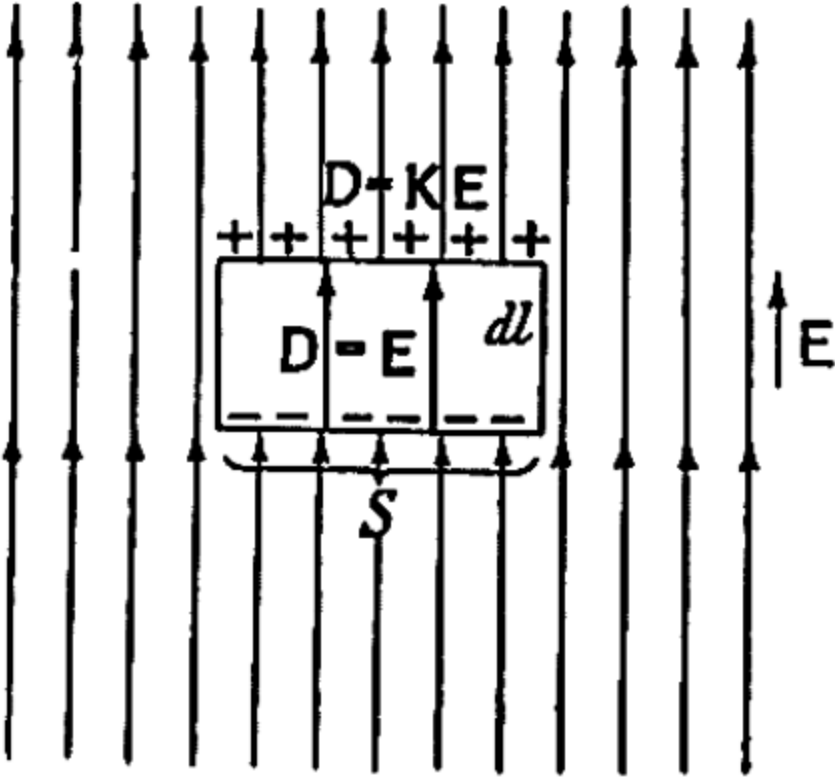


Fig. 3

Since  $E$  is to have the same value at all points, lines of displacement originate or terminate on the ends of the cylinder. Thus the ends carry true charges whose surface density may be calculated by equation (2) (p. 277). On the upper surface in fig. 3

$$\sigma_+ = \frac{1}{4\pi} (K |E| - |E|) = \frac{K-1}{4\pi} |E|,$$

while on the lower surface,

$$\sigma_- = \frac{1}{4\pi} (|\mathbf{E}| - K|\mathbf{E}|) = -\frac{K-1}{4\pi} |\mathbf{E}|.$$

Hence the volume element  $d\tau = S dl$  which was removed has a dipole moment amounting to

$$d\mathbf{P} = \frac{K-1}{4\pi} \mathbf{E} S dl = \frac{K-1}{4\pi} \mathbf{E} d\tau. \quad \cdot \cdot (8)$$

The dipole moment per unit volume is called the *polarization*  $\mathbf{P}$ :

$$\mathbf{P} = \frac{K-1}{4\pi} \mathbf{E} = \kappa \mathbf{E}, \quad \cdot \cdot \cdot \cdot (9)$$

and we thus have \*

$$\mathbf{D} = K\mathbf{E} = (1 + 4\pi\kappa)\mathbf{E} = \mathbf{E} + 4\pi\mathbf{P}. \quad \cdot \cdot \cdot (10)$$

The newly-introduced constant of the material,  $\kappa$ , which, by (9), stands in simple relation to  $K$ , is called the electrical *susceptibility* of the medium. In this manner we can replace all elements of volume of the dielectric by dipoles of moment  $\kappa \mathbf{E} d\tau$ .

The atomistic interpretation of the dielectric constant indicates that the vector  $\mathbf{P}$  is the truly physical quantity, while  $\mathbf{D}$  appears merely as an auxiliary mathematical quantity. As will be explained in detail later (p. 450), a resultant dipole moment arises in two ways in a volume element containing a large number of atoms or molecules. If the elementary particles are themselves dipoles (molecules with resultant positive and negative constituents), the random nature of their orientations makes the resultant moment zero, provided no external field acts. An applied field tends to bring the dipoles into its own direction, while thermal agitation tends to destroy this orderly arrangement. The greater the field, at a given temperature, the greater the resultant moment in the direction of the field. On the other hand, it is clear that for a given external field, this resultant will become smaller as the temperature is raised. This type of susceptibility, in analogy with the more familiar corresponding magnetic quantity, is called "para-electric". If, on the other hand, the elementary particles have normally no dipole moment, the field will produce one by distorting the electron shells. Of course this effect, which is independent of the temperature, occurs also in the case of para-electric susceptibility, but is masked by it on account of its smallness. If the entire dipole moment of the particle is thus produced by the applied field, we speak of "dia-electric" susceptibility.

If, in addition to the distribution of the true charges, we know the polarization as a function of position, then the electrostatic potential may be found by integration over the volume and surface elements. Since according to equation (28) (p. 273), the potential of a dipole of moment  $\mathbf{P} d\tau$  is given by  $V = (\mathbf{P}\mathbf{r}/r^3) d\tau$ ,

\* In the M.K.S. system  $\kappa = K - 1$  and  $\mathbf{D} = K_0\mathbf{E} + \mathbf{P}$ , but all tables of  $\kappa$  are given in terms of the definition (9).



the complete potential when dielectrics are present is given by

$$V = \int \frac{\rho d\tau}{r} + \int \frac{\sigma dS}{r} + \int \frac{\mathbf{P} \cdot \mathbf{r}}{r^3} d\tau. \quad . \quad . \quad . \quad (11)$$

The volume integral contributed by the dipoles of the polarized dielectric may be transformed into a surface integral as long as we are dealing with homogeneous bodies of uniform dielectric constant. Here  $\text{div } \mathbf{P}$  vanishes when no volume charges are present, according to equation (3) (p. 278) and equation (10) (p. 280). We therefore have

$$\oint \frac{\mathbf{P} d\mathbf{S}}{r} = \int \frac{1}{r} \text{div } \mathbf{P} d\tau + \int \mathbf{P} \text{grad}_s \frac{1}{r} d\tau = \int \frac{\mathbf{P} \cdot \mathbf{r}}{r^3} d\tau. \quad . \quad (12)$$

The field point is fixed during the integration, and the integration is extended over the several volume elements of the region. Conse-

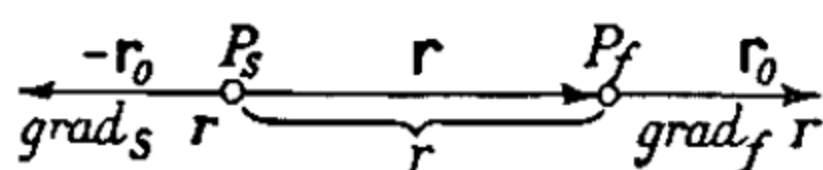


Fig. 4

quently, in taking the gradient, the field point is to be held fixed, while the "source points" are moved. This is indicated in (12) by the subscript  $s$ . According to

fig. 4, the gradient of  $r$  so obtained is equal and opposite to that obtained by moving the field point. We denote the former gradient by the subscript  $s$ , the latter by the subscript  $f$ . As before, the vector extending from the charge to the field point is denoted by  $\mathbf{r}$ .\*

The surface integral in equation (12) can also be given another interpretation: If an insulator is adjacent to a vacuum ( $\mathbf{P} = 0$ ), then

$$\mathbf{P} d\mathbf{S} = -\text{Div } \mathbf{P} dS,$$

and by equation (10) (p. 280), since  $\text{Div } \mathbf{D} = 0$ ,

$$\mathbf{P} d\mathbf{S} = \frac{\text{Div } \mathbf{E}}{4\pi} dS = \sigma' dS.$$

. . . . . (13)

Hence there are free charges on the surface of the dielectric as a result of the polarization there, and we may imagine the added field to have its origin in these charges, whose magnitude is given by equation (13) and equation (12).

\* In co-ordinate form it may be shown at once that each derivative of  $r$  with respect to the co-ordinates of the source point has a value equal and opposite to that of the derivative with respect to the co-ordinates of the field point. Thus, if the source point has the co-ordinates  $\xi, \eta, \zeta$ , and the field point has the co-ordinates  $x, y, z$ , we have

$$r = \sqrt{(x - \xi)^2 + (y - \eta)^2 + (z - \zeta)^2},$$

so that  $\partial r / \partial x = -\partial r / \partial \xi$ , &c.

### 3. Simple Examples of the Electrostatic Field in Dielectrics.

Equation (11) gives the potential, from which the field may be derived, if the distributions of the charges and of the polarization are known. But usually the polarization is no more known as a function of position than is the distribution of the charges (cf. p. 268). Since  $\mathbf{P} = \kappa \mathbf{E}$ , we must know the field strength  $\mathbf{E}$ , which has been altered on account of the presence of the dielectric. The problem admits of exact solution in only a few cases. In the absence of volume charges, we are concerned with finding a solution of  $\Delta V = 0$  which satisfies the conditions that (1)  $V$  has constant values on the surfaces of the conductors, and (2) at the surface of separation of two dielectrics,  $\text{Curl } \mathbf{E} = 0$  (continuity of the tangential components), and  $\text{Div}(K\mathbf{E}) = 0$  (prescribed discontinuity of the normal components of  $\mathbf{E}$ ). As an example of the problems permitting of exact solution we consider the distortion of an initially uniform field  $\mathbf{E}_0$  caused by the introduction of a dielectric sphere. We make the obvious assumption that the polarization *within* the sphere is homogeneous and parallel to  $\mathbf{E}_0$ , but we do not yet attempt to determine the magnitude of  $\mathbf{P}$ . We can show that, by suitable choice of the value of  $P$ , the added field arising from the dipoles of the polarized sphere, together with the original field, can be made to satisfy the boundary conditions. We may imagine the existence of the field of the uniformly polarized sphere to be due to a sphere homogeneously filled with charge of density  $-\rho$  and another filled with charge  $+\rho$ , the centres of the two being separated by  $ds$ . In order that every element of volume may have the dipole moment  $\mathbf{P} d\tau$ , we must have

$$\mathbf{P} d\tau = \rho d\tau ds, \quad . . . . . (14)$$

i.e. 
$$ds = \frac{\mathbf{P}}{\rho}.$$

Outside the sphere ( $r > a$ ) the added potential  $V_s$  is, by p. 261,

$$V_s^{(E)} = -\frac{4\pi a^3}{3r} \rho + \frac{4\pi a^3}{3} \rho \left( \frac{1}{r} + \frac{ds r}{r^3} \right) = \frac{4\pi a^3}{3} \frac{\mathbf{P} \mathbf{r}}{r^3}. \quad (15)$$

Since  $\text{grad}(\mathbf{P} \mathbf{r})$  is equal to  $\mathbf{P}$  when the vector  $\mathbf{P}$  is constant,\* the total field in the external space is

$$\begin{aligned} \mathbf{E}^{(E)} &= \mathbf{E}_0 - \frac{4\pi a^3}{3} \text{grad} \left( \frac{\mathbf{P} \mathbf{r}}{r^3} \right) \\ &= \mathbf{E}_0 - \frac{4\pi a^3}{3r^3} \mathbf{P} + \frac{4\pi a^3 \mathbf{P} \mathbf{r}}{r^5}. \quad . . . (16) \end{aligned}$$

\* This may be seen intuitively from the meaning of the gradient, or formally from  $\mathbf{P} \mathbf{r} = P_x x + P_y y + P_z z = u$  by calculating the components  $\partial u / \partial x$ , &c., or by taking the  $x$ -axis parallel to  $\mathbf{P}$ .

The field is therefore no longer constant in direction, for there is a component directed outward along the radius vector from the centre of the sphere (fig. 5).

Inside the sphere, the added potential is, by equation (25) (p. 272),

$$V_s^{(i)} = -2\pi\rho \left( a^2 - \frac{r^2}{3} \right) + 2\pi\rho \left( a^2 - \frac{r^2}{3} + \frac{2\mathbf{P}\mathbf{r}}{3\rho} \right) = \frac{4\pi}{3} \mathbf{P}\mathbf{r}, \quad (17)$$

$$\text{whence} \quad \mathbf{E}^{(i)} = \mathbf{E}_0 - \frac{4\pi}{3} \mathbf{P}. \quad (18)$$

Provided that our solution satisfies the boundary conditions, the internal field remains homogeneous, since  $\mathbf{P}$  is assumed to have the direction of  $\mathbf{E}$ . Thus we have in addition to the applied field  $\mathbf{E}_0$  another field of strength

$$\frac{4\pi}{3} \mathbf{P}$$

oppositely directed, whose source is the system of free surface charges connected with this field by equation (13) (p. 281):

$$\sigma' = -\text{Div } \mathbf{P}. \quad (19)$$

In the analogous magnetic case, one speaks of the demagnetizing effect of the boundary surfaces, and the numerical factor in (18)— $4\pi/3$  for the sphere—is called the demagnetization factor for this particular case.

We still have the magnitude of  $\mathbf{P}$  at our disposal in order to satisfy the boundary conditions. Denoting the angle between  $\mathbf{r}$  and  $\mathbf{E}_0$  (fig. 5) by  $\theta$ , we have for the tangential components

$$\mathbf{E}_{\text{tan}}^{(E)} = E_0 \sin \theta - \frac{4\pi a^3}{3a^3} P \sin \theta = \mathbf{E}_{\text{tan}}^{(i)} = E_0 \sin \theta - \frac{4\pi}{3} P \sin \theta. \quad (20)$$

The continuity of the tangential components is thus automatically fulfilled. For the normal components, the equation to be satisfied is

$$\begin{aligned} K_E \mathbf{E}_n^{(E)} &= K_E E_0 \cos \theta - K_E \frac{4\pi a^3}{3a^3} P \cos \theta + K_E \frac{4\pi a^3}{a^3} P \cos \theta \\ &= K_I \mathbf{E}_n^{(i)} = K_I E_0 \cos \theta - \frac{4\pi}{3} K_I P \cos \theta. \end{aligned} \quad (21)$$

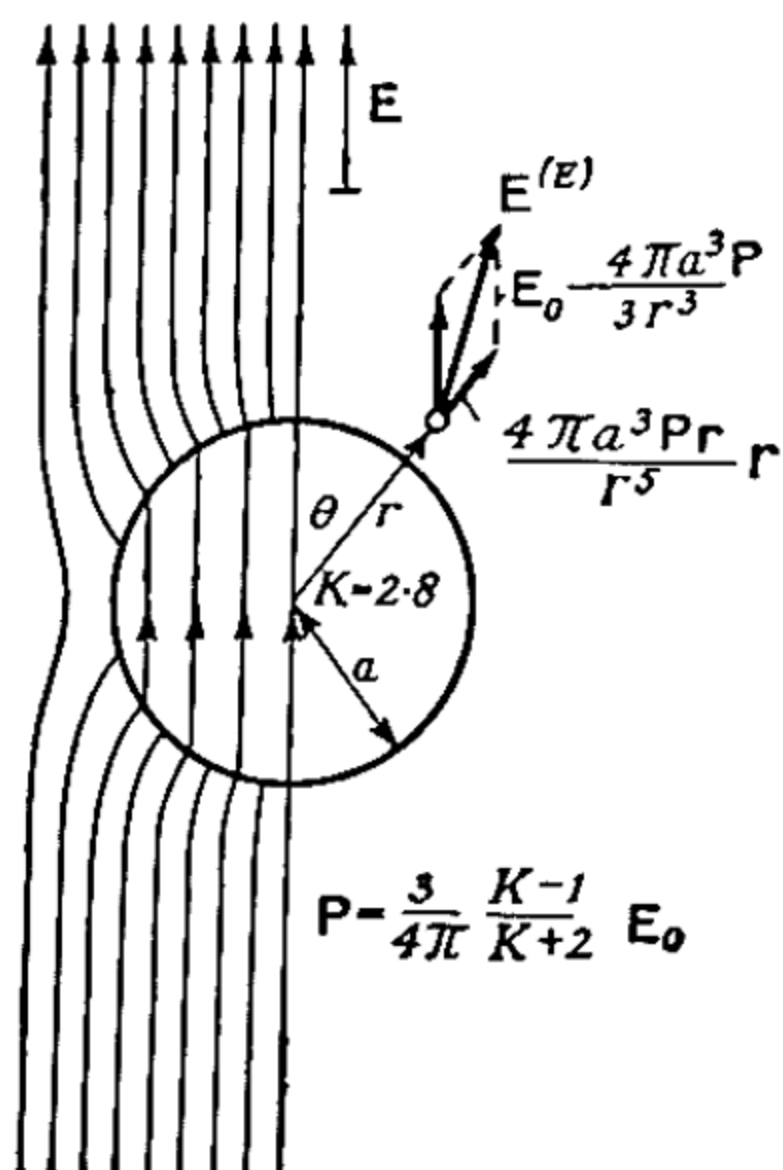


Fig. 5



For the present case, where  $K_e = 1$ ,  $K_i = K$ , this gives the following relation, which may be written at once in vector form since  $\mathbf{E}_0$  and  $\mathbf{P}$  have the same direction:

$$\mathbf{P} = \frac{3}{4\pi} \mathbf{E}_0 \frac{K-1}{K+2} \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad (22)$$

By substituting this value in equations (16) and (18), we obtain the field strengths within and exterior to the sphere. The former is

$$\mathbf{E}^{(i)} = \frac{3}{K+2} \mathbf{E}_0 \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad (23)$$

For infinitely large dielectric constant  $K_i$ ,  $\mathbf{E}^{(i)}$  vanishes. In electrostatics this is equivalent to the case of a conducting sphere. Hence putting  $K = \infty$  in equation (22) gives the polarization of an insulated conducting sphere placed in a uniform field  $\mathbf{E}_0$ :

$$\mathbf{P} = \frac{3}{4\pi} \mathbf{E}_0.$$

The total dipole moment of such a sphere is then

$$m = \frac{4\pi a^3}{3} \mathbf{P} = a^3 \mathbf{E}_0 \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad (24)$$

The polarizability, i.e. the ratio of dipole moment to field strength, is thus equal to the cube of the radius.

*Ex. 75.* In a uniformly polarized dielectric let the material within a sphere of radius  $a$  be removed without disturbing the polarization at any other place in the remaining medium. Determine the field produced at the centre of the sphere by the surrounding medium. (Observe that the supplementary field may be ascribed to free charges on the surface of the sphere.)

*Ex. 76.* Calculate the capacity of a spherical condenser whose dielectric has a specific inductive capacity  $K_1$  from  $r = a$  to  $r = b$  and one of magnitude  $K_2$  from  $r = b$  to  $r = c$  ( $a < b < c$ ).

*Ex. 77.* A point charge is near the infinite plane face of a dielectric medium. Calculate the field at points outside the dielectric and within it. (Here, as at p. 275, we can find a solution satisfying all the boundary conditions by placing a charge of suitable magnitude at the "image point" of the given charge.)

## CHAPTER XIII

### ENERGY AND PONDEROMOTIVE FORCES IN THE ELECTROSTATIC FIELD

#### 1. Potential Energy of Systems of Charges in a given Field.

In contrast with the procedure in the preceding section, we consider the electrostatic field  $\mathbf{E}$  to be given here, and introduce into it charges which are so small that they do not influence the distribution of the charges which produce the field.

The potential energy of a simple point charge at a place where the potential of the field is  $V$  is

$$u_e = eV. \quad . \quad . \quad . \quad . \quad . \quad . \quad (25)$$

If we have a dipole, and if the potential at the negative charge is  $V$ , the potential energy of the doublet is

$$u_d = -eV + e[V + ds \operatorname{grad} V] = -eds \mathbf{E} = -m\mathbf{E}. \quad (26)$$

In a homogeneous field the potential energy can change only by changing the angle between  $m$  and  $\mathbf{E}$ , and for pure translation the energy remains constant—i.e. a parallel displacement of a dipole in a uniform field involves no work. But since work is represented by the scalar product of force by displacement, this means that in a homogeneous field no force acts upon a dipole, tending to displace it. On the other hand, a rotation of amount  $\delta\theta$  requires an amount of work  $M\delta\theta$ , where  $M$  represents the magnitude of the torque acting upon the dipole. Hence

$$M\delta\theta = -\delta u_d = -E |m| \sin\theta \delta\theta,$$

and the turning moment acting on the dipole is therefore

$$M = - |m| E \sin\theta. \quad . \quad . \quad . \quad . \quad . \quad (27)$$

The negative sign means that the direction of the torque is that tending to decrease  $\theta$ .

So far we have considered the dipole moment to be constant, but the case where the dipole moment is proportional to the field, (i.e.

$m = \alpha E$ ) is equally important. A force  $eE$  acts upon the positive charge, tending to pull it away from the negative charge, considered fixed. At equilibrium, this force is balanced by an equal and opposite internal force. If the field increases by  $dE$ , the change in energy is equal to the change of energy of the dipole [equation (26)], plus the work done in displacing the charges by  $d\mathbf{r}$  against the internal forces. That is,

$$du_d = -m dE - E dm + eE d\mathbf{r} = -m dE = -\alpha E dE.$$

The total energy of the dipole induced by the field is obtained by integration from 0 to  $E$ :

$$u_d = -\frac{\alpha}{2} E^2 = -\frac{mE}{2}. \quad . \quad . \quad . \quad (28)$$

*Ex. 78.* Find the translational force acting upon a dipole of given orientation in a non-homogeneous electrostatic field.

## 2. Total Energy of the Electrostatic Field.

In electrostatics there are two different but essentially equivalent ways of considering the total energy of the field. The older method, which we shall call the theory of "action at a distance", considers only the charges and the forces operating between them, without attempting to explain how such action is brought about. On the other hand, the theory of direct action locates the energy in the intervening space also, and implies that wherever an electric field is present there is also a certain continuous distribution of energy. (The picture suggested is that of an elastic body, say the ether, in a state of stress. In spite of its great heuristic value, we can no longer ascribe physical reality to such a picture.) The superiority of the theory of direct action becomes apparent in the case of rapidly-varying fields, for it leads simply and naturally to the existence of electromagnetic waves.

We begin, however, with the theory of action at a distance, and consider the potential energy of two point charges  $e_1$  and  $e_2$  in a vacuum. This is given by the potential of the Coulomb force, and so may be written down at once:

$$u = \frac{e_1 e_2}{r_{12}}, \text{ or (M.K.S.) } u = \frac{e_1 e_2}{4\pi K_0 r_{12}}. \quad . \quad . \quad . \quad (29)$$

If there are other point charges, we can calculate the total energy by starting with the charge 1, bringing in charge 2 from infinity to the distance  $r_{12}$ , then bringing up charge 3 from infinity until it is at prescribed distances  $r_{13}$  and  $r_{23}$  from these charges, and so on. The total work done in thus building up the field—i.e. the potential energy—is



$$\begin{aligned}
 U &= \left( \frac{e_1 e_2}{r_{12}} + \frac{e_1 e_3}{r_{13}} + \frac{e_1 e_4}{r_{14}} + \dots \right) + \left( \frac{e_2 e_3}{r_{23}} + \frac{e_2 e_4}{r_{24}} + \dots \right) + \left( \frac{e_3 e_4}{r_{34}} + \dots \right) \\
 &= \frac{1}{2} \sum_i' \sum_k' \frac{e_i e_k}{r_{ik}}, \text{ or } \frac{1}{2} \sum_i' \sum_k' \frac{e_i e_k}{4\pi K_0 r_{ik}}. \quad \dots \quad (30)
 \end{aligned}$$

The factor  $1/2$  occurs before the double sum on account of the fact that every combination of  $i$  and  $k$  occurs twice in taking the double sum. The accents on the summation signs indicate that the combinations  $i = k$  are to be omitted. Now

$$\sum_i' \frac{e_i}{r_{ik}} \text{ or, in M.K.S. units, } \sum_i' \frac{e_i}{4\pi K_0 r_{ik}}$$

signifies the electrostatic potential, at the position occupied by the charge  $k$ , caused by all the other charges. Thus we may write

$$U = \frac{1}{2} \sum V_k e_k \text{ (same in M.K.S. units).} \quad \dots \quad (31)$$

The idea may be extended to include continuously distributed charges. For a volume charge,  $e = \rho d\tau$ : for a surface charge,  $e = \sigma dS$ . Hence the energy when volume and surface charges are present is

$$U = \frac{1}{2} \int V \rho d\tau + \frac{1}{2} \int V \sigma dS \text{ (same in M.K.S. units).} \quad (31')$$

In the theory of direct action we must ascribe a given energy content to every element of volume of the region where an electric field exists. The most natural assumption is that the *energy density*  $u$  (the energy per unit volume of the field) is proportional to the square of the electric field strength, for  $\mathbf{E}^2$  is the simplest scalar with which the scalar  $u$  may be connected. It is easy to verify, also, that  $\mathbf{E}^2$  has the dimension of energy per cubic centimetre in the electrostatic system. The factor of proportionality is to be determined in such a way that numerical agreement with the theory of action at a distance results. The equivalence of the expression so determined to (31') is easily shown. Thus, if we apply Gauss's theorem to the product  $\mathbf{E}V$ , and if we remember that  $\mathbf{E}$  vanishes as  $1/r^2$  and  $V$  vanishes as  $1/r$  at infinity, the integral over an infinite sphere is zero, so that

$$0 = \oint_{\infty} (\mathbf{E}V) d\mathbf{S} = \int \mathbf{E} \text{ grad } V d\tau + \int V \text{ div } \mathbf{E} d\tau. \quad \dots \quad (32)$$

Since  $\mathbf{E} = -\text{grad } V$  and  $\text{div } \mathbf{E} = 4\pi\rho$ , or  $= \rho/K_0$ ,

we have 
$$\int \mathbf{E}^2 d\tau = \int 4\pi V \rho d\tau, \text{ or } = \frac{1}{K_0} \int V \rho d\tau. \quad \dots \quad (33)$$

On the other hand, if only volume charges are present, the total energy of the field is, by (31'),

$$U = \frac{1}{2} \int V \rho d\tau.$$

We obtain numerical agreement if we take the energy density in a vacuum to be

$$u = \frac{1}{8\pi} \mathbf{E}^2 \text{ (C.G.S.)}, \text{ or } u = \frac{1}{2} K_0 \mathbf{E}^2 \text{ (M.K.S.)}. \quad (34)$$

If surface charges—and hence surfaces of discontinuity of  $\mathbf{E}$ —are present, these are to be excluded from the field by closely-fitting closed surfaces, when applying Gauss's theorem. This is the origin of the second integral in equation (31').

In a medium where the dielectric constant is  $K$ , the force between two point charges is decreased in the ratio  $1 : K$ , and the theory of action at a distance gives

$$U = \frac{1}{2} \sum_i' \sum_k' \frac{e_i e_k}{K r_{ik}} = \frac{1}{2} K \sum_i' \sum_k' \frac{e_i}{K} \frac{e_k}{K} \frac{1}{r_{ik}}. \quad (35)$$

But  $e_k/K$  is the apparent charge  $e_k'$ , from which the field in the medium of dielectric constant  $K$  may be derived, just as the field in a vacuum may be derived from the true charge  $e$ . Passing to a continuous distribution of charge, we write

$$U = \frac{K}{2} \int V \rho' d\tau + \frac{K}{2} \int V \sigma' dS. \quad (35')$$

If we again take  $u$  proportional to  $\mathbf{E}^2$  in the direct action method, the same transformation as was used in equation (32) gives

$$u = \frac{1}{8\pi} K \mathbf{E}^2 = \frac{1}{8\pi} \mathbf{E} \mathbf{D}, \text{ or } u = \frac{K K_0}{2} \mathbf{E}^2. \quad (36)$$

In the second form, the equation is also valid for anisotropic substances, where  $\mathbf{E}$  and  $\mathbf{D}$  have different directions.

We shall next consider the change in the energy of a field which results from introducing a dielectric body. The body will be assumed to be so small that its introduction does not alter the distribution of the charges producing the field. First let there be given a vacuum field  $\mathbf{E}_0 = \mathbf{D}_0$ . By bringing in the dielectric body, we produce a new field  $\mathbf{E}$  and a new displacement  $\mathbf{D}$  at every point. Hence the energy change is given by

$$U - U_0 = \frac{1}{8\pi} \int (\mathbf{E} \mathbf{D} - \mathbf{E}_0 \mathbf{D}_0) d\tau. \quad (37)$$

This expression may be rewritten as follows:

$$\mathbf{E}\mathbf{D} - \mathbf{E}_0\mathbf{D}_0 = \mathbf{E}(\mathbf{D} - \mathbf{D}_0) + (\mathbf{E} - \mathbf{E}_0)\mathbf{D}_0.$$

It is a fundamental property of the electrostatic field that  $\mathbf{E}$  is irrotational throughout the entire region (p. 267). Therefore we may set  $\mathbf{E}$  equal to  $-\text{grad } V$ . Again,  $\text{div}(\mathbf{D} - \mathbf{D}_0)$  is everywhere zero, and hence  $\mathbf{D} - \mathbf{D}_0$  is free of sources ("solenoidal"), since the true charges—the sources of  $\mathbf{D}$ —were not changed by the introduction of the dielectric substance. Moreover,  $\mathbf{E}$  and  $\mathbf{D}$  vanish at infinity as  $1/r^2$ , and  $V$  vanishes as  $1/r$ . Hence Gauss's theorem gives for integration over an infinite sphere:

$$0 = \oint_{\infty} V(\mathbf{D} - \mathbf{D}_0) d\mathbf{S} = - \int \mathbf{E}(\mathbf{D} - \mathbf{D}_0) d\tau + \int V \text{div}(\mathbf{D} - \mathbf{D}_0) d\tau.$$

Since  $\text{div}(\mathbf{D} - \mathbf{D}_0)$  vanishes everywhere, it follows that

$$\int \mathbf{E}(\mathbf{D} - \mathbf{D}_0) d\tau = 0$$

and

$$U - U_0 = \frac{1}{8\pi} \int (\mathbf{E} - \mathbf{E}_0) \mathbf{D}_0 d\tau. \quad . . . (38)$$

But for the same reason, the integral  $\int \mathbf{E}_0(\mathbf{D} - \mathbf{D}_0) d\tau$  vanishes also. If we multiply this integral by  $-1/8\pi$  and add it to (38), we obtain

$$\begin{aligned} U - U_0 &= \frac{1}{8\pi} \int (\mathbf{E}\mathbf{D}_0 - \mathbf{E}_0\mathbf{D}) d\tau = \frac{1}{8\pi} \int (\mathbf{E}\mathbf{E}_0 - \mathbf{E}_0\mathbf{D}) d\tau \\ &= -\frac{1}{2} \int \frac{K-1}{4\pi} \mathbf{E}\mathbf{E}_0 d\tau, \quad (39) \end{aligned}$$

since  $\mathbf{D}_0 = \mathbf{E}_0$  by hypothesis. From the definitions of electrical susceptibility and polarization (p. 280),

$$U - U_0 = -\frac{1}{2} \int \kappa \mathbf{E}\mathbf{E}_0 d\tau = -\frac{1}{2} \int \mathbf{P}\mathbf{E}_0 d\tau. \quad . . (40)$$

If the susceptibility is small,  $\mathbf{E}\mathbf{E}_0$  may be replaced by  $\mathbf{E}^2$  without introducing appreciable error; i.e. the alteration of the field may be neglected. If we make this substitution in the first integral of (40), we recognize at once the meaning of this expression. In a field  $\mathbf{E}_0$ , every cubic centimetre of the dielectric has a dipole moment  $\kappa\mathbf{E}_0$ , and by § 1 (p. 286) the energy of a dipole induced by the field is given by

$$-\frac{\kappa}{2} \mathbf{E}_0^2.$$

The change in the field within the dielectric, as calculated for the



sphere on p. 282, causes the scalar product  $\mathbf{E}\mathbf{E}_0$  to appear in place of  $\mathbf{E}_0^2$ .

This *decrease* of the field energy for *constant* field-producing charges, easily verified in the case of a parallel-plate condenser with and without dielectric, must not be confused with the *increase* in the energy content of a dielectric when the field is applied. In the latter situation the field-producing charges increase. According to (36) the change in energy density is

$$du = \frac{K}{4\pi} \mathbf{E} d\mathbf{E} = \frac{1 + 4\pi\kappa}{4\pi} \mathbf{E} d\mathbf{E} = \frac{1}{4\pi} \mathbf{E} d\mathbf{E} + \mathbf{P} d\mathbf{E}. \quad (36')$$

The last form shows that the second term represents the change in energy of a material body, in which  $\mathbf{P} \neq 0$ , while the first term represents the change in energy of a vacuum.

*Ex. 79.* Show that the energy of a charged condenser is  $\frac{1}{2}CV^2$  and compute the change in energy when a dielectric is introduced—in one case with the charge on the plates held fixed, in the other case with the potential difference held constant.

*Ex. 80.* Compute the work required to put together infinitely dispersed charges to form a sphere having a uniform charge density  $\rho$  and a radius  $a$ .

### 3. Forces and Equilibrium in the Electrostatic Field; Theory of the Manometer Method for the Electrical Susceptibility of a Liquid.

In mechanics, the position of equilibrium of a system may be determined by the condition that the potential energy must be a minimum. Now, according to § 1 (p. 285), the energy of the electrostatic field is equivalent to the potential energy of the charges, and

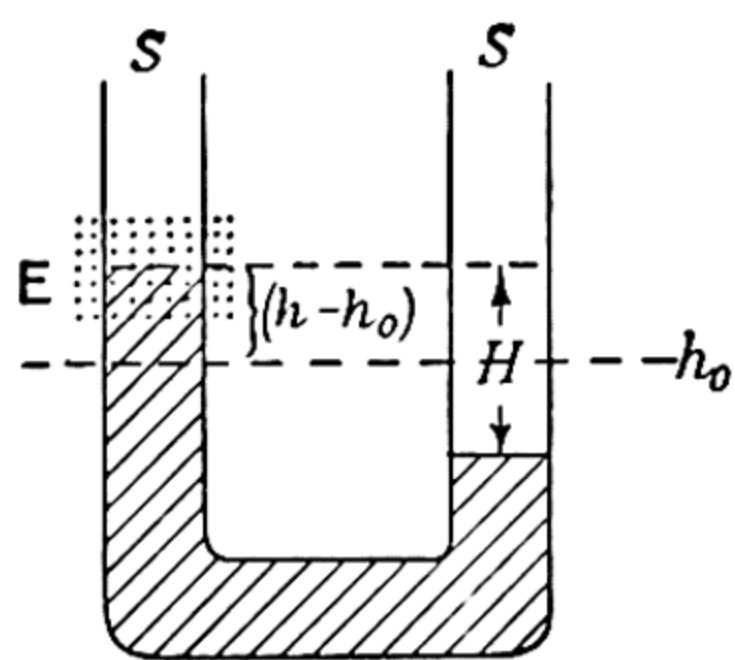


Fig 1.

so equilibrium is determined by the condition that *the sum of the electrical field energy and the mechanical potential energy* (e.g. that of gravity) *is to be a minimum*. We illustrate this principle with the theory of the manometer method for determining the electrical susceptibility of a liquid, which is especially important in the analogous magnetic case. A U-tube of cross-section  $S$  is filled with a dielectric liquid. Let an electric field, normal to the axis of the tube, be applied in the region of the meniscus in one arm of the tube (fig. 1). The lines of force are normal to the plane of the diagram, their trace in that plane being indicated by dots. Application of the field causes the meniscus in this side of the tube to rise. If the density of the liquid is  $\rho$ , and if the equilibrium height of the meniscus in the absence of the field is  $h_0$ , the work done against gravity amounts to  $S\rho(h - h_0)^2g$ ,

for the mass  $S\rho(h - h_0)$  has been moved from the right side to the left, whereby the centre of gravity has been raised a distance  $h - h_0$ . If the susceptibility is small, so that  $\mathbf{E}$  may be set equal to  $\mathbf{E}_0$ , the total energy becomes [cf. equation (40) (p. 289)]:

$$U = U_0 - \frac{1}{2} \int_0^h \kappa \mathbf{E}_0^2 S dh + S\rho(h - h_0)^2 g. \quad . \quad . \quad (41)$$

This means that the total energy depends only upon  $h$ ; hence the minimum condition gives

$$0 = \frac{dU}{dh} = -\frac{1}{2} \kappa [\mathbf{E}_0^2]_h S + 2S\rho(h - h_0)g, \quad . \quad . \quad (42)$$

and the difference in level of the menisci is

$$H = 2(h - h_0) = \frac{\kappa \mathbf{E}_0^2}{2\rho g}. \quad . \quad . \quad . \quad (43)$$

As is evident from the analysis,  $\mathbf{E}_0$  signifies the strength of the field applied to the meniscus on the left; the field need not be measured at other places.

In many cases it is not so important to know the position of equilibrium as it is to find the force actually operating on the system. For example, we may wish to find the motion of an uncharged oil drop in a non-homogeneous field, and must therefore know the forces acting. The droplet is assumed to be so small that the field may be considered homogeneous within the spherical volume (radius  $a$ ) which it occupies. Then, by equation (22) (p. 284), the polarization within the sphere is

$$\mathbf{P} = \frac{3}{4\pi} \frac{K - 1}{K + 2} \mathbf{E}_0,$$

and by the exact formula (40) (p. 289), the electrostatic energy is

$$U = U_0 - \frac{1}{2} \int \mathbf{P} \mathbf{E}_0 d\tau = U_0 - \frac{1}{2} \frac{4\pi}{3} a^3 \mathbf{P} \mathbf{E}_0 = U_0 - \frac{a^3}{2} \frac{K - 1}{K + 2} \mathbf{E}_0^2. \quad (44)$$

Since we assume the droplet to be so small that we may reckon with a constant value of  $\mathbf{E}_0$  within a volume of this magnitude, the energy of the system will depend on the position of the centre of the sphere. If the field  $\mathbf{E}_0$  is given, there is a definite value of the energy corresponding to each position of the centre of the droplet. Points at which this value is the same may be connected by level surfaces. If the sphere undergoes a displacement  $d\mathbf{s}$ , the change in the potential energy of the field is

$$dU = (\text{grad } U) d\mathbf{s} = - \frac{a^3}{2} \frac{K - 1}{K + 2} (\text{grad } \mathbf{E}_0^2) d\mathbf{s}. \quad . \quad (45)$$

On the other hand, the force acting on the sphere does work

$$dW = \mathbf{F} d\mathbf{s}.$$

Since this work must be equal to the *decrease* in potential energy, we have

$$\mathbf{F} = \frac{a^3}{2} \frac{K - 1}{K + 2} \text{grad } \mathbf{E}_0^2. \quad . \quad . \quad . \quad . \quad (46)$$

Since  $K > 1$ , the direction of the force is that in which  $\mathbf{E}_0^2$  increases; hence the droplet is urged toward regions of greater field strength. In magnetism, where we shall find far-reaching analogies with electrostatic phenomena (Chap. XV), the magnetic permeability  $\mu$  is the counterpart of  $K$ . We encounter values of  $\mu$  less than 1 as well as greater than 1, depending upon the nature of the substance. Spheres of those substances for which  $\mu < 1$  will be urged out of the magnetic field, and this difference in behaviour has led to the classification of magnetic materials into paramagnetic and diamagnetic substances.

*Ex. 81. The Absolute Electrometer.*—A flat metal plate of area  $S$  is suspended from one of the two equal arms of a balance so that it is at a distance  $a$  from another horizontal plate of the same size. What weight must be added to the other side of the balance to maintain equilibrium if the two plates are charged to a difference of potential  $V$  when the medium between the plates is (a) air, and (b) a material of dielectric constant  $K$ ?



## CHAPTER XIV

### STATIONARY ELECTRIC FIELDS AND STEADY CURRENTS

#### 1. Ohm's Law.

We know from what has gone before that it is impossible to maintain an electrostatic field in a conductor. However, by continuously supplying energy, a field which is constant in time can be maintained. This field differs from the electrostatic field in that, in the first place, it is accompanied by a continuous evolution of heat. But there is another difference: an actual transport of electric charge takes place in a field of this kind, as may be verified directly, for example, in the case of electrolytes. This continued motion of electric charges, which we shall refer to as an electric *current*, is also made evident by the magnetostatic field which accompanies it; this will be discussed in the following chapter.

The *current strength*, or simply the *current*  $I$ , is defined as the amount of electricity passing through a cross-section of the conductor in unit time:

$$I = \frac{de}{dt} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

The *current density*  $\mathbf{i}$  is a vector in the direction of motion of the electricity, whose magnitude is equal to the amount of electricity passing per second through unit area of cross-section normal to  $\mathbf{i}$ . Then the current through any section is given by

$$I = \int \mathbf{i} d\mathbf{S} = \int |\mathbf{i}| \cos(\mathbf{i}, \mathbf{n}) dS. \quad . \quad . \quad . \quad . \quad (2)$$

It has been found experimentally that for metals and electrolytes the current is proportional to the applied difference of potential (or E.M.F.)  $V$ . We write  $1/R$  for the factor of proportionality, and call  $R$  the *resistance* of the conductor.  $R$  may depend upon the temperature; hence *Ohm's Law*,

$$I = \frac{V}{R}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

assumes that precautions are taken to hold the temperature constant

in spite of the evolution of heat accompanying the passage of the current. In many instances the form

$$V = IR \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

has more direct physical meaning. This way of writing the equation states that if we have any conductor carrying a current, the potential difference between two points between which there is a resistance  $R$  is  $IR$ . If we have a closed, unbranched circuit consisting of various resistances, the first of Kirchhoff's laws of electric circuits says that the algebraic sum of the potential differences across the conductors is equal to the sum of the electromotive forces (batteries, generators, &c.) applied in that circuit. The other law of Kirchhoff states that at any junction, the sum of all inflowing currents must be equal to the sum of all outflowing currents. This is merely a special case of the absence of sources or sinks (solenoidal property) for electric currents, and is almost self-evident.

For the theory of direct action we require a formulation of Ohm's Law which connects only the quantities localized at one particular point. For a cylindrical conductor the resistance  $R$  is proportional to the length and inversely proportional to the area of cross-section. The resistance offered by a centimetre cube of material to a current passing between opposite faces is called the *specific resistance* or *resistivity*  $\rho$  of the substance; its reciprocal,  $\sigma$ , is called the *conductivity*. Hence the expression for the current in a cylindrical conductor may be written

$$I = \frac{VS}{l} \sigma, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

Now  $V/l$  is equal to the field in the direction of the generators of the cylinder and  $I/S$  is the current density. Since the current is in the direction of the field for isotropic bodies, these considerations hold for any small cylinder drawn within a conductor carrying a current. Hence (5) may be written in the vector form

$$\mathbf{i} = \sigma \mathbf{E}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

The fundamental implication of Ohm's Law is that, for conductors such as metals and electrolytes to which it applies,  $\sigma$  is independent of the field strength.

The electric field strength is an irrotational vector in current-carrying conductors, just as it is in electrostatics. Also, there are no surface vortices on the surface of separation of two conductors:

$$\text{Curl } \mathbf{E} = [\boldsymbol{\nu}(\mathbf{E}_2 - \mathbf{E}_1)] = 0 \quad \text{or} \quad \mathbf{E}_{\tan}^{(2)} = \mathbf{E}_{\tan}^{(1)}. \quad . \quad (7)$$

But the normal component of  $\mathbf{E}$  experiences a discontinuity at the interface; the surface divergence of  $\mathbf{i}$  must vanish, since any charge

which enters the surface from one side must emerge from the other, so that

$$\text{Div } i = n(i_2 - i_1) = n(\sigma_2 E_2 - \sigma_1 E_1) = 0 \text{ or } \sigma_2 E_n^{(2)} = \sigma_1 E_n^{(1)}. \quad (8)$$

Thus the change in direction of the lines of flow at the surface of separation of two conductors is the same as that of the electrostatic lines of force at the surface separating two dielectrics (p. 279):

$$\frac{\tan \alpha_1}{\tan \alpha_2} = \frac{\sigma_1}{\sigma_2}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (9)$$

Likewise, no lines of flow can originate within the conductor, so that we have

$$\text{div } i = \text{div}(\sigma E) = -\sigma \text{div grad } V = -\sigma \Delta V = 0. \quad . \quad (10)$$

This means that inside homogeneous conductors  $V$  satisfies the same differential equation as it does in the electrostatic field in free space. Moreover, the boundary conditions are also the same, except that the constants have a different meaning. This property of steady fields may be used to map the field experimentally. If it is desired to determine the static field about a body of arbitrary form, this can be done by immersing a metallic model in a large metal basin filled with an electrolyte. The body is connected to one pole of a source of potential, and the basin to the other. With the help of a Wheatstone bridge the equipotential surfaces may be determined with ease. Since the conductivity of the solution is small compared with that of the metal, the condition that the surface of the body be an equipotential surface is practically realized.

*Ex. 82.* A condenser of capacity  $C$  is at the potential  $V_0$  at time  $t = 0$ . If it is discharged through a resistance  $R$ , calculate the resulting current.

*Ex. 83.* A condenser  $C$  is charged through a large resistance  $R$  by means of a source of direct potential  $V_0$ . A glow lamp having a striking potential  $V_s$  is connected in parallel with the condenser. What is the frequency of the resulting pulsating current?

## 2. Generation of Heat in a steady Electric Field.

According to p. 267, the work done when an electric charge  $e$  is moved from a point where the potential is  $V_1$  to a point at potential  $V_2$  is

$$W = e(V_1 - V_2). \quad . \quad . \quad . \quad . \quad . \quad . \quad (11)$$

If the carrier of the charge is free to move, this work appears as kinetic energy; in steady fields within conductors, the work is transformed into heat. The atomistic view of conduction interprets this heat as that arising from "frictional" motion of the charges. Hence, in a section of a conductor bounded by two equipotential surfaces at



potentials  $V_1$  and  $V_2$ , the amount of heat generated per second by a current  $I$  is

$$W = I(V_1 - V_2), \quad . \quad . \quad . \quad . \quad . \quad (12)$$

measured in energy units, since  $I$  is the quantity of electricity which passes through a potential difference  $V_1 - V_2$  each second. Using Ohm's law to eliminate the potential difference from (12), we have the usual form of the equation for the so-called Joule heat loss,

$$W = I^2 R. \quad . \quad . \quad . \quad . \quad . \quad (12a)$$

From the view-point of the theory of direct action, a differential formulation is desirable. Let us take a volume element in the form of a small cylinder whose lateral surface is formed by lines of flow and whose ends are portions of equipotential surfaces. If the altitude of the cylinder is  $h$  and the cross-section is  $S$ , then the potential difference between the end surfaces is

$$V_1 - V_2 = h |\text{grad } V| = h |\mathbf{E}| \quad (V_1 > V_2).$$

The current is  $I = S |\mathbf{i}|$ , so that the quantity of energy dissipated as heat in the volume element  $d\tau$  each second is

$$dW = |\mathbf{i}| |\mathbf{E}| Sh = \sigma \mathbf{E}^2 d\tau = \frac{\mathbf{i}^2}{\sigma} d\tau. \quad . \quad . \quad . \quad (13)$$

It may be shown without difficulty that this equation is true for an element of volume of any kind.

## CHAPTER XV

### THE MAGNETOSTATIC FIELD

#### 1. Comparison of Electrostatic and Magnetostatic Fields.

Besides electrical phenomena, there is a second type of non-mechanical effects whose close connexion with electricity was recognized at an early period, viz. *magnetic* phenomena. Just as an electrostatic field can be maintained without energy supply, so there exist also magnetic fields which persist without the necessity of supplying energy. The field between the poles of a permanent magnet is of this type, also the magnetic field which, as we shall see, accompanies a steady current. These are included in the classification *magnetostatic field*. It is true that energy is dissipated continuously in the latter case, but this is merely a consequence of the finite resistance of the conductor, and has nothing to do with the magnetic field, whose strength depends solely upon the current. If the resistance of a conductor could be made arbitrarily small—a condition which is almost realized at very low temperatures (cf. p. 449)—these magnetostatic fields could also be maintained without consuming energy.

If, in analogy with the electrostatic field, we try to map the field by using a magnetic test body, we encounter the first difference between electrostatic and magnetostatic fields: no test body exists having an independent magnetic charge. Even the smallest fragment of a steel magnet is a dipole which behaves in a magnetic field much as an electrical dipole does in an electrostatic field. Hence we can explore the field only by observing the torque exerted upon a small magnet of known magnetic moment  $m$ . From the calculation for the corresponding electrical case (p. 285), this torque is

$$|M| = -|m| |H| \sin \theta, \text{ or in vector form, } \mathbf{M} = [m\mathbf{H}], \quad (1)$$

where  $\mathbf{H}$  is the magnetic field strength. The position of equilibrium is that for which  $\theta = 0$ ; for small displacements, the twisting couple, or "moment of torsion", is

$$\tau = |m| |H|. \quad \dots \dots \dots (1')$$

If the moment of inertia of the small bar magnet is known,  $\tau$  may be determined experimentally by observing the period of vibration in the field [use equation (20) (p. 148)]. This is the well-known Gauss method for determining the horizontal intensity of the earth's magnetic

field. In order to obtain the field strength  $\mathbf{H}$  from the value of the torque  $M$ , we must also know the moment  $m$  of the test magnet. Since the fields of magnetic systems follow the same laws as those of electrical charges, we can calculate the field of a magnetic doublet by the results of p. 272; then, by observing the forces experienced by other dipoles in this field, we can obtain a second equation connecting the two unknown quantities  $m$  and  $\mathbf{H}$ . A detailed description of such an experiment may be found in any standard treatise on electricity and magnetism.\* The absence of isolated magnetic charges, or poles, implies

$$\operatorname{div} \mathbf{H} = 0 \quad . \quad . \quad . \quad . \quad . \quad (2)$$

at all points in a vacuum. *The magnetostatic field in a vacuum is everywhere solenoidal* (free of sources, or non-divergent). The electrostatic field, which may have a finite divergence even in a vacuum, is nevertheless characterized by complete absence of curl. This is by no means always the case for a magnetostatic field, as may be seen from the simple example of the field about a straight wire carrying a current. As we know from elementary experiments, the lines of force in this case are circles whose centres are at the wire, their planes being normal to the wire. If we take the line integral  $\oint \mathbf{H} d\mathbf{s}$  along one of these circles it cannot vanish, since we are continually adding quantities of the same algebraic sign. Hence the irrotational condition is no longer fulfilled by the magnetic field, for paths encircling lines of flow. Thus, *while the electrostatic field in a vacuum must be everywhere irrotational but may be divergent, the magnetostatic field in a vacuum is non-divergent at all points, but may have a finite curl.*

In the same way as the electrostatic field between the plates of a condenser is reduced by a factor  $K$  when a medium of dielectric constant  $K$  replaces the vacuum between the plates, so the field between the poles of a magnet is changed by the introduction of a medium having different magnetic properties. The quantity corresponding to the dielectric constant  $K$ , which is analogously defined as the ratio  $\mathbf{H}_{\text{vac}} : \mathbf{H}$ , is called the magnetic permeability  $\mu$ . The quantity  $\mu\mathbf{H}$  corresponding to the dielectric displacement  $\mathbf{D}$  is called the magnetic induction  $\mathbf{B}$ . As in the electrical case, the effect of the material may be represented by a polarization of the medium. The magnetic quantity corresponding to the electrical magnitude  $\mathbf{P}$  (the polarization) is called the *intensity of magnetization*

$$\mathbf{M} = \kappa\mathbf{H} = [(\mu - 1)/4\pi]\mathbf{H}, \quad . \quad . \quad . \quad . \quad . \quad (3)$$

where  $\kappa$  is the magnetic susceptibility. The intensity of magnetization

\* If the force is measured in C.G.S. units,  $m$  and  $\mathbf{H}$  will thus be given in this system. The unit of magnetic field strength so defined was formerly called the *gauss* but is now designated the *oersted*. In the M.K.S. system the field is specified in terms of the current producing it (cf. § 2).



is also equal to the magnetic moment per unit volume. The magnetic susceptibility differs in an important way from the electric susceptibility: there are substances for which  $\kappa$  is negative as well as substances for which it is positive. In contrast with electric polarization, which is always in the same direction as the field, the magnetization may be either in the direction of the field or in the opposite direction. But even in the electrical case, we found (p. 280) two different ways in which the polarization may come about. The same two possibilities exist in the magnetic case also, except that the difference is far more apparent here, inasmuch as the one kind (alignment of the already existing dipoles) yields positive values of  $\kappa$  (paramagnetic substances), while the other kind (creation of dipoles by induction) gives negative values of  $\kappa$  (diamagnetic substances). While the electrical susceptibility of liquids and solids lies between 0.1 and 10, the magnetic susceptibility, with few exceptions, amounts only to about  $10^{-4}$  or even less. But for the so-called *ferromagnetic* metals, which are mostly metals of the iron group and certain alloys,  $\kappa$  assumes enormously greater values (of the order of  $10^3$ ); however,  $\kappa$  is not constant, but depends upon the strength of the applied field. This fact may be understood from the way in which paramagnetic susceptibility is believed to arise: if the resultant dipole moment may be ascribed to the alignment of existing dipoles, then this moment cannot remain proportional to  $\mathbf{H}$  while  $\mathbf{H}$  increases indefinitely, since an upper limit is attained when all the elementary dipoles finally have been brought into line with the field.

After this digression on the subject of the atomistic interpretation of susceptibility we return to the description of the experimental properties of the magnetostatic field, which is wholly analogous to that for the electrostatic field. On account of the absence of true magnetic charges, we have always

$$\operatorname{div} \mathbf{B} = \operatorname{div} (\mu \mathbf{H}) = 0 \quad \text{and} \quad \operatorname{Div} \mathbf{B} = \mu_1 \mathbf{H}_n^{(1)} - \mu_2 \mathbf{H}_n^{(2)} = 0. \quad (4)$$

On the other hand,  $\mathbf{H}$  is derived from apparent (free) charges residing in non-homogeneous parts of the media:

$$\operatorname{div} \mathbf{H} = 4\pi\rho_m, \quad \operatorname{Div} \mathbf{H} = 4\pi\sigma_m. \quad . \quad . \quad . \quad (5)$$

If no finite ("surface") current flows in the surface of separation of two media—a case which we exclude—then  $\operatorname{Curl} \mathbf{H} = 0$ . Hence the refraction of the lines of force at the interface is governed by the tangent law

$$\frac{\tan \alpha_1}{\tan \alpha_2} = \frac{\mu_1}{\mu_2}. \quad . \quad . \quad . \quad . \quad (6)$$

Just as there are no true single magnetic charges, so there exist no conductors of magnetism. On account of the high value of  $\mu$  for

ferromagnetic substances, the lines of force reaching these are very nearly normal to the surface, as may be seen from (6). This means that such substances behave very much like conductors, with regard to the static arrangement of the lines of force.

The expression for the magnetic energy density also corresponds to that for the electrostatic case:

$$u_{\text{mag}} = \frac{\mu}{8\pi} \mathbf{H}^2 = \frac{1}{8\pi} \mathbf{B}\mathbf{H}, \text{ or } = \frac{\mu_0 \mu \mathbf{H}^2}{2} \text{ in M.K.S. units; }^* \quad (7)$$

hence the calculations of § 3, Chap. XIII (p. 290), may be applied to magnetism, as already mentioned there.

*Ex. 84.* As a first approximation, the earth may be regarded as a sphere magnetized with uniform intensity. Calculate the angle of dip as a function of the magnetic latitude.

*Ex. 85.* In a certain terrestrial magnetic survey, an area of  $13.1 \times 10^6 \text{ km.}^2$  was enclosed by a path from New York to England and back. The path was described in the counter-clockwise direction, as seen from above. The result was that  $\oint \mathbf{H} ds = -64 \times 10^4 \text{ oersted cm.}$  What does this mean?

## 2. Calculation of the Magnetostatic Field accompanying a Given Distribution of Electric Currents in a Vacuum.

In elementary physics the magnetic field due to steady currents is calculated by means of a formula from the theory of action at a distance. This law, which gives the magnetic field due to an element of a conductor, is the familiar rule of Biot and Savart (p. 303). However, this law is not a suitable basis for obtaining a more general relationship corresponding to the direct-action point of view; moreover, the Biot-Savart rule cannot be verified experimentally, since there are no isolated elements of conductors—at least for steady fields—but only closed circuits. Hence we start from a different empirical fact. If we map the magnetic field of a conducting wire, as outlined on p. 297, and calculate the line integral  $\oint \mathbf{H} ds$  along any arbitrary curve which encircles the conductor, we obtain the same value, proportional to the current in the circuit, no matter how we choose the path of integration, as long as the circuit threads it:

$$\oint \mathbf{H} ds = \gamma I. \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

If we measure  $\mathbf{H}$  in Gaussian magnetic units [the connexion is given directly by the Gauss method of determining the field (p. 298)], and if  $I$  is measured in the electrostatic system, then, in order to make the equation numerically correct, we must take the value

$$\gamma = 4\pi/c,$$

\* The quantity  $\mu_0$  will be defined on p. 312.

where  $c$  is numerically equal to the velocity of light in a vacuum, expressed in cm./sec. If, on the other hand, we take  $\gamma = 4\pi$ , we thereby determine a new unit for  $I$ —the *electromagnetic C.G.S. unit*, which, however, we wish in principle to avoid using in spite of the fact that the practical units such as the ampère and the volt differ from the electromagnetic C.G.S. units only by powers of ten. In the M.K.S. system there is no specific unit of magnetic field strength, but the field is defined by setting  $\gamma = 1$  in equation (8). Thus the field is specified by the current producing it, and has the dimension amp./metre.

We wish to bring equation (8) into a form consonant with the theory of direct action. Imagine that we have given an extended conductor in which the current density vector  $\mathbf{i}$  is known as a function of position. If we draw a small closed curve  $C$  inside this conductor, the current passing through it is

$$I = \mathbf{i} \cdot d\mathbf{S},$$

where  $d\mathbf{S}$  denotes the directed element of area bounded by  $C$ . From the definition of curl,

$$\oint_C \mathbf{H} \cdot d\mathbf{s} = (\text{curl } \mathbf{H}) \cdot d\mathbf{S}$$

for this element of area. According to equation (8), this must be equal to

$$\frac{4\pi I}{c} = \frac{4\pi \mathbf{i} \cdot d\mathbf{S}}{c}.$$

Since this relationship must hold for any element of area whatsoever, for slowly varying fields

$$\text{curl } \mathbf{H} = \frac{4\pi \mathbf{i}}{c} = \frac{4\pi}{c} \sigma \mathbf{E}. \quad . \quad . \quad . \quad . \quad (9)$$

or in M.K.S. units,  $\text{curl } \mathbf{H} = \sigma \mathbf{E}. \quad . \quad . \quad . \quad . \quad . \quad (9')$

This is the differential form of the law corresponding to the theory of direct action. The vector equation states that the field lines of the current density vector are the same as those of the vector  $\text{curl } \mathbf{H}$ . Since the divergence of a curl is identically zero (p. 28), the lines of flow can neither originate nor terminate in the finite region. They must either form closed curves or go to infinity.

The problem for the electrostatic field was to deduce an *irrotational* field from a given distribution of sources, or charges. This task was simplified by the introduction of the electrostatic potential, which could be computed readily by integration over all the charges. In the complementary problem now before us—the calculation of



*solenoidal* or *non-divergent* field from the distribution of the vortices (currents)—we likewise simplify our task by introducing an auxiliary quantity. This quantity is no longer a scalar, but a vector, which we call the *vector potential*. We take

$$\mathbf{H} = \text{curl } \mathbf{V}. \quad . . . . . (10)$$

The non-divergence of  $\mathbf{H}$  is thus provided for. Then by equation (98) (p. 40),

$$\text{curl } \mathbf{H} = \text{curl curl } \mathbf{V} = \text{grad div } \mathbf{V} - \Delta \mathbf{V} = \frac{4\pi \mathbf{i}}{c}. \quad . (11)$$

Without loss of generality, we can assume that  $\text{div } \mathbf{V}$  vanishes; for if it did not do so, we might add to  $\mathbf{V}$  an irrotational vector having the same sources as  $\mathbf{V}$ , but with opposite signs, without changing  $\mathbf{H}$  in any way. Thus we obtain the same type of equation for the vector potential as for the scalar potential in electrostatics:

$$\Delta \mathbf{V} = -\frac{4\pi \mathbf{i}}{c}. \quad . . . . . (12)$$

In component form (cf. p. 40):

$$\Delta V_x = -\frac{4\pi i_x}{c}, \quad \Delta V_y = -\frac{4\pi i_y}{c}, \quad \Delta V_z = -\frac{4\pi i_z}{c}. \quad (13)$$

The solution of this set of equations is known from electrostatics: if we are dealing only with currents distributed in space (surface currents, corresponding to surface charges in electrostatics, are temporarily excluded), the solution is

$$V_x = \int \frac{i_x d\tau}{cr}, \quad V_y = \int \frac{i_y d\tau}{cr}, \quad V_z = \int \frac{i_z d\tau}{cr}. \quad . (14)$$

The component forms may be combined again into a vector equation:

$$\mathbf{V} = \int \frac{\mathbf{i} d\tau}{cr}. \quad . . . . . (15)$$

By means of this equation the magnetic field of a linear conductor (wire) may be determined at once. Here the vector  $\mathbf{i}$  is in the direction of the axis. Let the element of length  $ds$  be in this direction also, and let the cross-sectional area of the wire be  $S$ . Then

$$\mathbf{i} d\tau = \frac{I}{S} S ds, \quad . . . . . (16)$$

and thus, since  $I$  has the same value at all points of the conductor,

$$\mathbf{V} = \frac{I}{c} \int \frac{ds}{r}. \quad . . . . . (17)$$

The integration is to be extended over all the elements of the conductor and has nothing to do with taking the curl of  $\mathbf{V}$ , in which operation the co-ordinates of the field point are differentiated. Hence

$$\mathbf{H} = \frac{I}{c} \text{curl} \int \frac{d\mathbf{s}}{r} = \int \frac{I}{c} \text{curl} \frac{d\mathbf{s}}{r} = \int \frac{I}{c} \left[ \text{grad} \frac{1}{r}, d\mathbf{s} \right] = \int \frac{I}{c} \frac{[d\mathbf{s}, \mathbf{r}_0]}{r^2}. \quad (18)$$

The factor  $4\pi/c$  on the right side of (12) is absent in the M.K.S. system, so that in these units

$$\mathbf{H}_{\text{M.K.S.}} = \int \frac{I}{4\pi} \frac{[d\mathbf{s}, \mathbf{r}_0]}{r^2}. \quad \cdot \cdot \cdot \cdot \cdot \quad (18')$$

Here  $\mathbf{r}_0$  is a unit vector drawn from the segment  $d\mathbf{s}$  toward the field point. We can interpret this equation to mean that each element of length of the wire contributes an amount

$$\frac{I}{c} \frac{[d\mathbf{s}, \mathbf{r}_0]}{r^2}$$

to the field. But this is precisely the *Law of Biot and Savart*, according to which the field at  $P$  due to the current in the element  $d\mathbf{s}$  is normal to the plane determined by  $d\mathbf{s}$  and the radius vector from  $d\mathbf{s}$  to  $P$ , and is of amount  $I ds \sin \theta / cr^2$ .

Again, it may be shown without much difficulty that a closed linear current (circuit) gives rise to the same magnetic field as an arbitrary surface covered by magnetic dipoles which have the directions of the normals to this surface, and which have a magnetic moment of  $I/c$  per unit area, the surface being bounded by the circuit. A surface thus uniformly covered by normally set magnetic doublets is called a *uniform, normally magnetized shell*. It is, of course, further understood that the poles of one kind are all on one side of the shell. This equivalence of a magnetic shell and a current may easily be proved with the help of certain transformations of vector analysis. We must be careful to note whether the field point or the element of the conductor is the element which varies. All differentiation referring to a displacement of the field point will be characterized by the subscript  $f$ ; derivatives relating to variation of the element of the conductor will be denoted by the subscript  $s$ . We again note, as on p. 281, that

$$\text{grad}_f f(r) = -\text{grad}_s f(r).$$

First we transform the expression for the vector potential at the point  $P$  by a generalization of Stokes's theorem. Let  $u$  be a scalar point function, and let it be required to transform the line integral  $\oint u d\mathbf{s}$  into a surface integral. We multiply scalarly by an arbitrary constant vector  $\alpha$  and obtain the following result by using Stokes's theorem and equation (89) (p. 38):

$$\begin{aligned} \alpha \oint u d\mathbf{s} &= \oint u \alpha d\mathbf{s} = \int [\text{curl}(u\alpha)] d\mathbf{S} = \\ &= \int [\text{grad } u, \alpha] d\mathbf{S} = \alpha \int [d\mathbf{S}, \text{grad } u]. \end{aligned}$$

Since  $\alpha$  is perfectly arbitrary, we must have in general

$$\oint u d\mathbf{s} = \int [d\mathbf{S}, \text{grad } u].$$

If we apply this equation to the vector potential as given by equation (17) (p. 302), we must remember that the field point is fixed, and hence all derivatives relate to change of the elements of the conductor. Hence

$$\mathbf{V} = \frac{I}{c} \oint \frac{d\mathbf{s}}{r} = \frac{I}{c} \int \left[ d\mathbf{S}, \text{grad}_s \frac{1}{r} \right] = \frac{I}{c} \int \left[ \mathbf{n}, \text{grad}_s \frac{1}{r} \right] dS. \quad (19)$$

If now we compute the field strength  $\mathbf{H}$  by taking the curl, the field point is to vary, so that

$$\begin{aligned} \mathbf{H} &= \frac{I}{c} \text{curl}_r \int \left[ \mathbf{n}, \text{grad}_s \frac{1}{r} \right] dS = \frac{I}{c} \int \text{curl}_r \left[ \mathbf{n}, \text{grad}_s \frac{1}{r} \right] dS \\ &= - \frac{I}{c} \int \text{curl}_r \left[ \mathbf{n}, \text{grad}_r \frac{1}{r} \right] dS. \end{aligned}$$

The integrand may be transformed by introducing the  $\nabla$ -operator:

$$\begin{aligned} \text{curl}_r \left[ \mathbf{n}, \text{grad}_r \frac{1}{r} \right] &= \left[ \nabla_r \left[ \mathbf{n} \nabla_r \frac{1}{r} \right] \right] \\ &= \mathbf{n} \cdot \nabla_r \nabla_r \frac{1}{r} - \nabla_r \cdot \mathbf{n} \nabla_r \frac{1}{r}. \quad (20) \end{aligned}$$

Since, however,  $\nabla_r \nabla_r \frac{1}{r} = \Delta_r \frac{1}{r} = 0$ , we obtain

$$\begin{aligned} \mathbf{H} &= + \int \text{grad}_r \left( + \frac{I}{c} \mathbf{n} \cdot \text{grad}_r \frac{1}{r} \right) dS \\ &= - \text{grad}_r \int \left( \frac{I}{c} \frac{\mathbf{n} \cdot \mathbf{r}}{r^3} \right) dS. \quad (21) \end{aligned}$$

Hence the magnetic field is derivable from the scalar potential

$$V_m = \int \frac{I}{c} \frac{\mathbf{n} \cdot \mathbf{r}}{r^3} dS = \frac{I}{c} \int \frac{\cos(\mathbf{n} \cdot \mathbf{r})}{r^2} dS, \quad (22)$$

where  $\mathbf{r}$  is the radius vector drawn to the field point from the element  $dS$  of any surface bounded by the conductor. Comparison with the



expression for the potential of a double layer [(29'), p. 273] shows that the dipole moment per unit area is  $I/c$  and that the direction of the dipoles is that of the normal  $\mathbf{n}$  to the surface, the sense being such that the current flows in the clockwise direction when sighting along  $\mathbf{n}$ . In M.K.S. units the dipole moment per  $\text{cm.}^2$  is  $\mu_0 I$ .

But the potential has different values on the two sides of the surface, since the integral  $\oint \mathbf{H} d\mathbf{s}$  must amount to  $4\pi I/c$  for a complete circuit, and the contribution of the arbitrarily small space occupied by the shell must vanish. This confirms the result found on p. 274, that the potential has a discontinuity amounting to  $4\pi |\mathbf{m}|$  at a double layer. The result of (21) may be stated: *At points outside any surface bounded by a closed circuit, the magnetic field is that which would be produced by a double layer covering this surface uniformly.\** The surface may have any form. It represents a surface of discontinuity such that on passing through it the potential must be increased by  $4\pi |\mathbf{m}|$ .

*Ex. 86.* Calculate the magnetic field at interior and exterior points of an infinitely long solid cylindrical conductor of radius  $R$ .

### 3. Calculation of the Magnetic Field accompanying Electric Currents when Ferromagnetic Materials are present.

If we fill the region about a current-carrying conductor with bodies of various permeabilities, experiment shows that the line integral  $\oint \mathbf{H} d\mathbf{s}$  taken over any closed curve continues to have the value  $4\pi I/c$ , where  $I$  is the total current enclosed by the path of integration. It follows that *equation (9) (p. 301) is independent of the permeability of the medium.* Hence, in contrast with electrostatics, where the field due to charged bodies is reduced by a factor  $K$  by filling the space with a medium of dielectric constant  $K$ , the *magnetic field* of currents remains unaltered if the entire space be filled with a medium of permeability  $\mu$ . This is true, however, only provided the entire region is thus filled (see below).

Since, on the other hand, the formal application of the theory developed for electrostatics requires that the magnetic field of fictitious single charges or dipoles be reduced by a factor  $\mu$  in a medium of permeability  $\mu$ , we must increase the strength of the equivalent magnetic shell of the circuit by this factor if  $\mathbf{H}$  is to remain the same when the permeability is changed to  $\mu$ . Thus we must give the shell a moment  $\mu I/c$  per unit area, or in M.K.S. units,  $\mu\mu_0 I$ .

If, however, we place ferromagnetic materials in certain portions

\* This was stated by Ampère in his *Théorie des phénomènes électro-dynamiques, Mémoires de l'Institut*, IV (1823), as follows: "Every linear conductor carrying a current is equivalent to a simple magnetic shell, the bounding edge of which coincides with the conductor, and whose moment per unit area—that is, the strength of the shell—is proportional to the strength of the current."

of the region only, a change in the field occurs in the magnetic case also, and this change is considerable on account of the high value of the permeability. Since equation (9) (p. 301) is independent of the medium, the field  $\mathbf{H}$  and the field  $\mathbf{H}_0$  in a vacuum must have the same vortex lines, viz. the lines of flow of the current, which are the field lines of the vector  $\mathbf{i}$ . Hence the field to be added to  $\mathbf{H}_0$ , viz.  $\mathbf{H} - \mathbf{H}_0$ , must be irrotational and hence derivable from a scalar potential. It naturally suggests itself that this scalar potential may be simply that of the dipoles of the magnetized bodies. This is really the case, and may be demonstrated as follows: Let us assume the usual case where we deal with individual ferromagnetic bodies, each of uniform permeability. Then by § 1 (p. 299), there exists no volume divergence of  $\mathbf{H}$ , but there is a *surface* divergence at the bounding surfaces. This divergence is connected with the apparent surface charges whose surface density is  $\sigma_m'$  by the relation

$$\text{Div } \mathbf{H} = \text{Div } (\mathbf{H} - \mathbf{H}_0) = 4\pi\sigma_m'.$$

On the other hand, on account of the absence of spatial source, (charges),

$$\text{div } \mathbf{H} = \text{div } (\mathbf{H} - \mathbf{H}_0) = -\text{div grad } V = -\Delta V = 0.$$

The solution of this equation is known from electrostatics. If the only sources are those distributed on surfaces, this solution is

$$V = \int \frac{\sigma_m'}{r} dS.$$

Now since  $\text{Div } \mathbf{B}$  is always zero (absence of true magnetic charges), we have, from (3), p. 298,

$$\text{Div } \mathbf{M} = -\frac{1}{4\pi} \text{Div } \mathbf{H} = -\sigma_m'.$$

Hence the potential, from which the supplementary field is derived becomes

$$V = -\int \frac{\text{Div } \mathbf{M}}{r} dS. \quad . \quad . \quad . \quad . \quad (23)$$

The integration is to be extended over the surfaces of all the ferromagnetic bodies. Since these must be separate closed surfaces, we have the added potential in the form given for the electrostatic case at the end of § 2, Chap. XII (p. 281). Using Gauss's theorem, and reversing the argument used there, we can represent the potential as an integral over the volumes occupied by ferromagnetic substances:

$$V = -\Sigma \oint \frac{\text{Div } \mathbf{M}}{r} dS = \Sigma \oint \frac{\mathbf{M} d\mathbf{S}}{r} = \Sigma \int \frac{\mathbf{M} \boldsymbol{\nu}}{r^3} d\tau. \quad (24)$$



But the last expression is exactly the potential of the dipoles of the magnetized bodies; this is the fact we proposed to demonstrate.

If the field which produces the magnetization  $\mathbf{M}$  were not itself altered by the presence of the ferromagnetic substances, calculation of the field would be a simple matter. In that case, one could first compute the vacuum field strength  $\mathbf{H}_0$  from the distribution of the currents, then determine from this the magnetization of the bodies introduced, and thus calculate the supplementary field of the dipoles from this result. How far from the truth this result would be is shown by the example of the sphere in an originally homogeneous field—a problem which we already treated for the electrostatic case. According to equation (23) on p. 284, the interior field which determines the magnetization is no longer  $\mathbf{H}_0$ , but

$$\mathbf{H}^{(i)} = \frac{3}{\mu + 2} \mathbf{H}_0, \quad . . . . . (25)$$

so that for  $\mu = 2000$ ,  $\mathbf{H}^{(i)}$  is only about 0.15 per cent of  $\mathbf{H}_0$ .

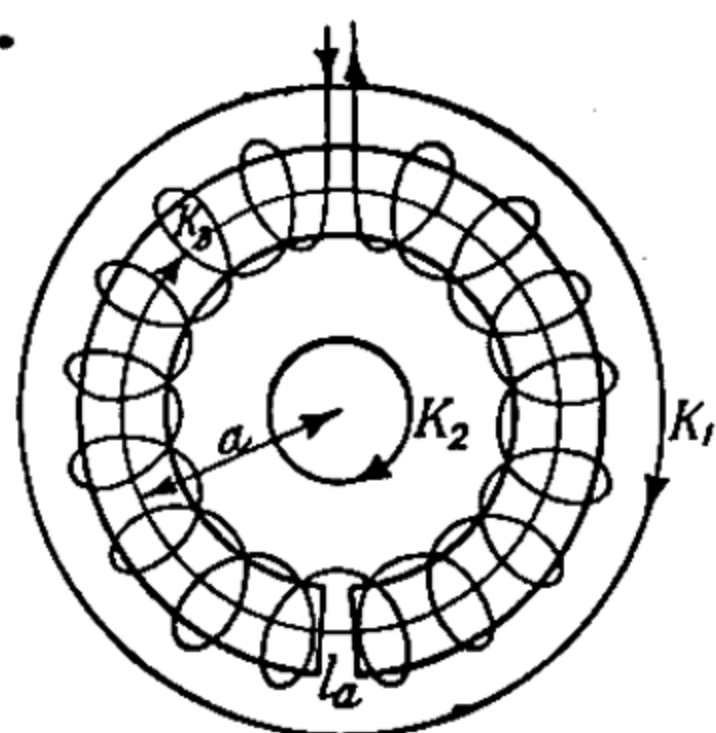


Fig. 1

As we saw above in the electrostatic case, the exact calculation of the field is possible only for bodies of simple form. A case of practical importance, viz. the field between the poles of a ring-shaped electromagnet, will be calculated *approximately* in another way. We choose a special kind of ring-magnet consisting of a ring or tore of ferromagnetic material spirally wound uniformly over its whole length with a wire carrying a current (fig. 1). From the symmetry of the arrangement we see that the magnetic lines of force must be circles

whose centres are on the axis of symmetry of the ring, and which lie in planes perpendicular to that axis. If we take the integral  $\oint \mathbf{H} d\mathbf{s}$  along a circle  $K_1$  or  $K_2$  lying entirely outside the ring, we see that this integral is zero, since no lines of flow are threaded by the path of integration. On the other hand, for such a circle within the ring, each of the  $N$  turns is threaded once, so that the integral over the entire path is  $4\pi NI/c$ . Thus the field is confined to the space within the windings and is of amount

$$H = 4\pi NI/(2\pi ac) = 2NI/(ac), \text{ or } H = NI/(2\pi a), \quad (26)$$

where  $a$  is an average radius. The M.K.S. measure of the field is especially vivid, for  $NI/(2\pi a)$  is merely the number of ampere-turns per metre. If, as we shall assume, the cross-section of the ring is small compared with its diameter, the field within the ring will be approxi-





1. The moving-coil galvanometer. A cylindrical coil of  $N$  turns is suspended in a uniform magnetic field by a torsion fibre which is a continuation of a diameter of the coil (fig. 2). If a current  $I$  flows in the coil, the magnetic moment of the coil in air is  $NIS/c$ , where  $S$  is the area enclosed by each turn. The magnetic moment vector is normal to the plane of the coil. If this moment makes an angle  $\theta$  with the field  $\mathbf{H}$ , then by equation (27) (p. 285) the torque acting on the suspended system is

$$M = -\frac{NIS}{c} H \sin \theta. \quad . . . . . (29)$$

This moment tends to turn the coil into the stable position where  $\mathbf{M}$  is parallel to  $\mathbf{H}$ . The torsion of the suspension opposes this turning. If  $\theta = \pi/2$  is the rest position when no current flows, this torque is a maximum. If the displacements are small, we can take  $\sin \theta = 1$ . Since the restoring moment of the fibre is proportional to the displacement  $\phi = \pi/2 - \theta$ , the equilibrium position is determined by

$$\tau \phi = \frac{NSH}{c} I, \quad . . . (29')$$

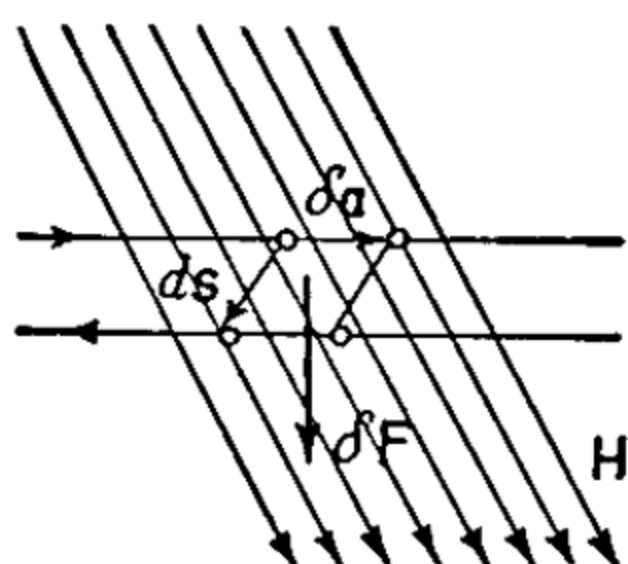


Fig. 3

where  $\tau$  is the moment of torsion of the fibre.

2. Force on a segment of a current-carrying conductor in a magnetic field. Imagine a segment  $ds$  of a circuit which may be slid back and forth (fig. 3). We seek to determine the force which the uniform field  $\mathbf{H}$  exerts on this segment. By Ampère's theorem and (26), p. 285, the energy of the entire circuit is, in air ( $\mu \approx 1$ ):

$$U = -\frac{I}{c} \int \mathbf{H} d\mathbf{S} = -\frac{I}{c} \mathbf{H} \int d\mathbf{S} = -\frac{I}{c} \mathbf{H} S. \quad . (30)$$

If the area is slightly changed by giving  $ds$  a displacement  $\delta \mathbf{a}$ , the energy change is

$$\delta U = -\frac{I}{c} \mathbf{H} \delta \mathbf{S}.$$

The vector  $\delta \mathbf{S}$  may be represented by the vector product  $[\delta \mathbf{a} d\mathbf{s}]$ . If the factors are taken in this order, the normal is obtained in the right direction.\* This displacement is accompanied by an amount of work  $\mathbf{F} \delta \mathbf{a}$  done by the force. This work must be equal to the decrease in energy of the field:

$$\mathbf{F} \delta \mathbf{a} = -\delta U = +\frac{I}{c} \mathbf{H} [\delta \mathbf{a} d\mathbf{s}] = +\frac{I}{c} [d\mathbf{s} \mathbf{H}] \delta \mathbf{a},$$

\* The current in the segment, the magnetic field, and the ponderomotive force must form a right-handed orthogonal system, in this order (Fleming's left-hand rule).

so that the force itself is

$$\mathbf{F} = \frac{I}{c} [ds \mathbf{H}] = \frac{1}{c} [i \mathbf{H}] d\tau, \quad . . . . (31)$$

where  $d\tau$  is the volume of the part of the conductor of length  $ds$ . This force, tending to enlarge or decrease the area of the circuit (depending upon the direction of the current), is always present, whether it causes the conductor to move or not. If the current in a coil is very large, the magnetic field of the coil itself gives rise to large mechanical forces acting on the various portions of the winding.

By p. 305, the magnetic moment per unit area is increased by a factor  $\mu$  if a medium having a permeability of this amount replaces the air. In this case formulæ (30) and (31) must be multiplied by  $\mu$ , and we have for the force

$$\mathbf{F} = \frac{\mu}{c} [i \mathbf{H}] d\tau, \text{ or } \mathbf{F}_{\text{M.K.S.}} = \mu\mu_0 [i \mathbf{H}] d\tau. \quad . . (31')$$

Although in computing the force we may replace a circuit by a magnetic shell, we must avoid setting the potential energy used here equal to the field energy. The reason is that something new appears when the conductor is moved through the field, viz. an *induced electromotive force* (see next chapter) whose effect is to weaken the current. In order to maintain the former current, *energy must be supplied*. The net effect is that for a *fixed* current the total field energy is found to increase by exactly the amount by which the magnetostatic potential energy would—according to the above calculation—decrease.

*Ex. 87. The Ballistic Galvanometer.*—A current of very short duration is sent through the coil of a suspended-coil galvanometer at the time  $t = 0$ , the total quantity of electricity which flows being  $e$ . What is the maximum displacement? (Take the time integral of the force, and compute the amplitude of the vibration from the initial velocity thus found.)



## CHAPTER XVI

### SLOWLY VARYING (QUASI-STATIONARY) FIELDS

#### *Definition of quasi-stationary fields*

All electric and magnetic fields so far considered have been independent of the time. We now go a step further and consider fields which vary with the time, but restrict ourselves for the present to fields varying relatively slowly. These are of great importance. We shall find later that electrical fields are propagated with the velocity of light. If, then, we restrict ourselves to not too rapid changes, and to systems whose dimensions are not too great, we may assume that a state corresponding to instantaneous propagation holds at every point of the field. The relationships derived on this basis hold under all conditions for the range of alternating current frequencies in technical use (10 to 1000 cycles per second).

#### 1. The Law of Induction. Maxwell's Equations.

Imagine a closed linear conductor in a magnetic field. Let the surrounding medium have a permeability  $\mu$ . We know from elementary experimental physics that if the field  $\mathbf{H}$  changes, a current is "induced" in the conductor. This means that there must be an electric field  $\mathbf{E}$  within the conductor. Since we assume the conductor to be linear,  $\mathbf{E}$  has the direction of the linear elements  $d\mathbf{s}$ . If we compute the integral  $\oint \mathbf{E} d\mathbf{s}$ , say in the direction of  $\mathbf{E}$ , this integral can no longer vanish, since all the contributions to the sum are of one sign. We thus encounter a rotational electric field for the first time. The value of  $\oint \mathbf{E} d\mathbf{s}$  is called the induced electromotive force, for the following reason. If we imagine the conductor severed between two neighbouring points  $P_1$  and  $P_2$ , the work which must be done to move unit charge from  $P_2$  to  $P_1$  against the field is

$$-W = \oint \mathbf{E} d\mathbf{s},$$

since the contribution of the open portion  $P_1P_2$  is vanishingly small. Hence there is a difference of electrical potential\* of this amount between  $P_1$  and  $P_2$ . The Law of Induction states that the induced elec-

\* This may manifest itself, for example, in the form of a spark passing between the two points.

tromotive force is equal to  $1/c$  times the decrease per second of the number of lines of magnetic induction enclosed by the circuit.\* The usual orientation of surface normal and direction of describing the circuit is assumed. The mathematical formulation of the Law of Induction is then

$$\oint \mathbf{E} d\mathbf{s} = -\frac{1}{c} \frac{\partial}{\partial t} \int \mu \mathbf{H} d\mathbf{S}. \quad . . . . . (1)$$

In the M.K.S. form there is a quantity on the left having the dimension *volts*; on the right, *amp./sec.* Hence there appears a new dimensional constant, called the permeability of free space, whose numerical value is

$$\mu_0 = 1.256 \times 10^{-6} \text{ volt sec./amp.}$$

The M.K.S. form of the law is thus

$$\oint \mathbf{E} d\mathbf{s} = -\frac{\partial}{\partial t} \int \mu_0 \mu \mathbf{H} d\mathbf{S}. \quad . . . . . (1a)$$

The line integral  $\oint \mathbf{E} d\mathbf{s}$  may be changed to a surface integral by using Stokes's theorem:

$$\oint \mathbf{E} d\mathbf{s} = \int \text{curl } \mathbf{E} d\mathbf{S} = -\frac{1}{c} \frac{\partial}{\partial t} \int \mu \mathbf{H} d\mathbf{S}. \quad . . . (2)$$

For stationary objects, integration with respect to position and differentiation with respect to time are independent, so that we may write

$$\int \text{curl } \mathbf{E} d\mathbf{S} = -\frac{1}{c} \int \frac{\partial}{\partial t} (\mu \mathbf{H}) d\mathbf{S}. \quad . . . . . (3)$$

Since this equation must be satisfied for any surface, the integrands of the surface integrals must be equal, and we obtain the first triple set of *Maxwell's Equations* of the electromagnetic field:

$$\text{curl } \mathbf{E} = -\frac{\mu}{c} \frac{\partial \mathbf{H}}{\partial t} = -\frac{1}{c} \frac{\partial \mathbf{B}}{\partial t}. \quad . . . . . (3a)$$

In the M.K.S. system,

$$\text{curl } \mathbf{E} = -\mu \mu_0 \frac{\partial \mathbf{H}}{\partial t}. \quad . . . . . (3b)$$

The Maxwellian theory assumes that this induced electric field is present even if not detectable in the form of a potential difference between the ends of a material conductor; in other words, this set

\* This statement, sometimes referred to as Neumann's Law, is a fact of experience. The part giving the direction of the induced electromotive force is an example of what is known as Lenz's Law.

of equations is generally valid for expressing the magnitude of the electric field induced by a changing magnetic field.

*Ex. 88.* Compute the current induced in the coil of a moving-coil galvanometer as it swings, and find the damping of the motion of the coil produced by this current.

## 2. Self and Mutual Induction.

### (a) Calculation of the coefficient of mutual induction

Let there be given two circuits 1 and 2. If a current flows in circuit 1, it generates a magnetic field  $\mathbf{H}_1$  whose lines of force pass through circuit 2. If the current  $I_1$  in circuit 1 changes, an electromotive force is induced in circuit 2. Its magnitude is given by

$$V_2 = \oint \mathbf{E}_2 d\mathbf{s}_2 = -\frac{\mu}{c} \int \frac{\partial \mathbf{H}_1}{\partial t} d\mathbf{S}_2 = -\frac{\mu}{c} \frac{\partial}{\partial t} \int \mathbf{H}_1 d\mathbf{S}_2. \quad (4)$$

If the whole of space is filled with a medium of uniform permeability  $\mu$ , then according to Chap. XV, § 2 (p. 302),  $\mathbf{H}_1$  is derivable from a vector potential:

$$\mathbf{H}_1 = \text{curl } \mathbf{V}_1, \quad \mathbf{V}_1 = \frac{1}{c} \int \frac{i_1 d\tau_1}{r_{12}} = \frac{I_1}{c} \oint \frac{d\mathbf{s}_1}{r_{12}}.$$

Hence, by Stokes's Theorem,

$$\int \mathbf{H}_1 d\mathbf{S}_2 = \int (\text{curl } \mathbf{V}_1) d\mathbf{S}_2 = \oint_2 \mathbf{V}_1 d\mathbf{s}_2 = \frac{I_1}{c} \oint_1 \oint_2 \frac{d\mathbf{s}_1 d\mathbf{s}_2}{r_{12}}, \quad (5)$$

i.e. the electromotive force induced in circuit 2 becomes

$$V_2 = \oint \mathbf{E}_2 d\mathbf{s}_2 = -\frac{\mu}{c^2} \frac{\partial I_1}{\partial t} \oint \oint \frac{d\mathbf{s}_1 d\mathbf{s}_2}{r_{12}}. \quad \cdot \cdot \cdot (6)$$

The quantity  $\frac{\mu}{c^2} \oint \oint \frac{d\mathbf{s}_1 d\mathbf{s}_2}{r_{12}}$ , or  $\frac{\mu\mu_0}{4\pi} \oint \oint \frac{d\mathbf{s}_1 d\mathbf{s}_2}{r_{12}}$  M.K.S. units (henry)

which depends only upon the geometric relationships of the two circuits and upon the permeability of the intervening medium is called the *coefficient of mutual induction*, or the *mutual inductance*  $L_{12}$ . With this notation (6) may be written

$$V_2 = \oint \mathbf{E}_2 d\mathbf{s}_2 = -L_{12} \frac{\partial I_1}{\partial t}. \quad \cdot \cdot \cdot \cdot (6')$$

Since both circuits enter into the expression for  $L_{12}$  symmetrically, we obtain the same expression if we calculate the electromotive force induced in circuit 1 by a change in the current flowing in circuit 2, i.e.

$$L_{12} = L_{21}. \quad \cdot \cdot \cdot \cdot (7)$$



(b) *Self-induction*

Even if there is but one circuit, a change in current causes a change in the magnetic flux through the circuit which, by the Law of Induction, results in an induced electromotive force. In the same way as we defined the coefficient of mutual induction above, we define the *coefficient of self-induction*, or the *self-inductance* to be a certain positive magnitude  $L_{11}$  depending only upon the geometric form of the circuit and the permeability of the surrounding medium. The defining equation is

$$\oint \mathbf{E} d\mathbf{s} = - L_{11} \frac{dI}{dt} \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad (8)$$

Here, too, we must use the negative sign, since the direction of the induced electrical field is such as to oppose the original current (Lenz's Law). Were it the opposite way, the original current would be increased by the self-induction, hence a still greater electromotive force would be induced, and so on without limit.

If we attempt the calculation of the self-inductance of a circuit in a way similar to the above computation of mutual inductance—i.e. by calculating the field at  $d\mathbf{s}$  caused by all the other current elements, then computing the total induced electromotive force by integration over all conducting elements—the resulting formulæ are found to have only limited application. In calculating mutual inductance we are permitted to assume that the conductors are linear, since the separation of the two circuits is taken to be large compared with their cross-sections. For the case of self-inductance, however, this is no longer the case, and we must divide the current into individual filaments and take into account their mutual influence. Thus the calculation of self-inductance is generally more troublesome than that of mutual inductance, so that we shall omit the detailed treatment here.

In the case of coils, we are mainly concerned with the magnetic field at the position of a selected turn caused by the current in the remaining turns; hence a division of the current into filaments is not required. In this instance the coefficient of self-induction may be calculated simply and to a good approximation by another method. We start again with a ring-magnet of  $N$  turns, the cross-section  $S$  being small compared with the radius  $a$  of the ring. The field inside the windings is [equation (26) (p. 307)]

$$H = \frac{4\pi NI}{2\pi ac} = \frac{4\pi NI}{lc}.$$

If the strength of the current changes, the total magnetic flux

through the  $N$  turns will also change, and this change, per second, is \*

$$\frac{d}{dt} \int \mu \mathbf{H} d\mathbf{S} = \frac{d\Phi}{dt} = \mu NS \frac{dH}{dt} = \mu NS \frac{4\pi N}{lc} \frac{dI}{dt},$$

so that the induced electromotive force is

$$V = \oint \mathbf{E} d\mathbf{s} = - \frac{4\pi}{c^2} \frac{N^2 S \mu}{l} \frac{dI}{dt}.$$

The self-inductance of the ring-shaped coil is therefore

$$L_{11} = \frac{4\pi N^2 S \mu}{c^2 l}, \text{ or } \frac{\mu \mu_0 N^2 S}{l} \text{ in M.K.S. units.} \quad . \quad . \quad (9)$$

This result holds approximately for a very long unclosed coil also. The approximation consists in neglecting the disturbance of the uniform field caused by the ends of the coil.

(c) *Transformation of the expression for the magnetic field energy*

The energy of a system of current-bearing conductors is mainly that of the magnetic fields caused by the several currents. Since the electric fields present in such systems are small, the electrostatic energy is relatively negligible. With the help of the concepts of self and mutual induction, the expression for the magnetic energy may be brought into a form which is very useful for many purposes. According to equation (7) (p. 300),

$$U_m = \frac{1}{8\pi} \int \mu \mathbf{H}^2 d\tau. \quad . \quad . \quad . \quad (10)$$

We assume that space is filled with a uniform medium of permeability  $\mu$ , although the final expression obtained will be valid even if the permeability varies from place to place.  $\mu \mathbf{H}$  is non-divergent everywhere, but  $\mathbf{H}$  is non-divergent only in a homogeneous medium. We may thus write

$$\mathbf{H} = \text{curl } \mathbf{V}.$$

Now by equation (92) (p. 38),  $\mathbf{H}(\text{curl } \mathbf{V}) = \mathbf{V}(\text{curl } \mathbf{H}) + \text{div } [\mathbf{V}\mathbf{H}]$ . Substituting this expression in equation (10), we can transform the second integral into a surface integral over an infinite sphere. However, by p. 302 we see that  $\mathbf{V}$  decreases as  $1/r$ , so that  $\text{curl } \mathbf{V}$  decreases as  $1/r^2$ , and hence this integral vanishes. As a result, we have

$$U_m = \frac{\mu}{8\pi} \int \mathbf{V} (\text{curl } \mathbf{H}) d\tau = \frac{\mu}{2c} \int \mathbf{V} i d\tau. \quad . \quad . \quad (11)$$

\* Here, and in the formulæ following, integration implies summation over all  $N$  turns.

If we have only linear conductors, which, of course, must be closed if currents are to flow at all, (11) yields

$$U_m = \frac{\mu}{2c} \sum_k I_k \oint \mathbf{V} ds, \quad . \quad . \quad . \quad . \quad . \quad (11')$$

where  $I_k$  is the current in the  $k$ th conductor. If we now decompose the vector potential at the position of circuit  $k$  into the parts  $\mathbf{V}_1, \mathbf{V}_2, \dots$  contributed by the various conductors (including circuit  $k$ ), then we have by equation (15) (p. 300) and the definitions of the coefficients of induction,

$$\oint \mathbf{V}_i ds_k = \frac{c}{\mu} L_{ik} I_i \quad \text{and} \quad \oint \mathbf{V}_k ds_k = \frac{c}{\mu} L_{kk} I_k, \quad . \quad (12)$$

Hence we have

$$U_{\text{mag}} = \frac{1}{2} (L_{11} I_1^2 + 2L_{12} I_1 I_2 + \dots + 2L_{1n} I_1 I_n + L_{22} I_2^2 + 2L_{23} I_2 I_3 + \dots + L_{nn} I_n^2). \quad (13)$$

*Ex. 89.* A charge  $-e$  revolves about a fixed charge  $+e$  in a circle of radius  $a$ . How much energy is absorbed by this system when a magnetic field  $\mathbf{H}$  is applied, the angle between the field and the normal to the orbit being  $0$ ?

(To simplify the calculation, assume that  $\left| \frac{\partial \mathbf{H}}{\partial t} \right| = \text{const.}$ )

*Ex. 90.* (a) Calculate the mutual inductance of two coaxial circles of wire, radii  $a_1$  and  $a_2$ , whose centres are at a distance  $z$  apart.

(b) Show that for two conductors  $L_{11}L_{22} > L_{12}^2$  always.

### 3. Stationary Alternating Current Circuits.

With the help of complex numbers we can derive relationships for the stationary state of uniform periodicity attained by circuits to which periodically alternating potentials are applied. The steady state referred to is that arrived at some time after the initial application of the electromotive force. These relationships permit us to extend formally to alternating current circuits the laws of Ohm and of Kirchhoff derived above for direct currents.

Let us consider a circuit consisting of a source of periodically alternating electromotive force  $V$ —say a dynamo—connected across a conductor of ohmic (steady current) resistance  $R$  and self-inductance  $L$ . In addition to the applied E.M.F., there is also that due to the self-induction of the circuit, so that

$$V - L \frac{dI}{dt} = IR. \quad . \quad . \quad . \quad . \quad . \quad (14)$$

If  $R$  may be neglected, we have

$$V = L \frac{dI}{dt}. \quad . \quad . \quad . \quad . \quad . \quad (14')$$



Assuming the phenomenon to be strictly sinusoidal,

$$V = V_0 e^{i\omega t}, \quad I = I_0 e^{i\omega t}. \quad . \quad . \quad . \quad . \quad . \quad (15)$$

Hence

$$\frac{dI}{dt} = i\omega I_0 e^{i\omega t} = i\omega I,$$

and so

$$V = I i\omega L. \quad . \quad . \quad . \quad . \quad . \quad (16)$$

Comparing this expression with Ohm's law, we see that a formal correspondence is obtained by setting the alternating-current resistance caused by self-induction—the so-called *reactance* of the circuit—equal to  $i\omega L$ . What is the physical meaning of this imaginary resistance of self-induction? It has been found of great advantage to represent alternating current phenomena by means of the vector diagram (cf. p. 52). If we place the "vector" of the current along the positive real axis of such a diagram, (16) shows that the vector for the electromotive force lies in the positive imaginary axis. That is, in a circuit containing only an alternator and a pure inductance, the electromotive force *leads* the current by a phase angle  $\pi/2$ . For a given applied potential, the strength (amplitude) of the current is given by (16) as

$$I_0 = \frac{V_0}{L\omega}. \quad . \quad . \quad . \quad . \quad . \quad (17)$$

Suppose, now, that the inductive portion of the circuit is replaced by a capacity (condenser)  $C$ . If the applied E.M.F. were constant, then after the initial charging current no current would flow, but on account of the periodic charging and discharging of the condenser, an alternating current results. Here again we obtain an imaginary resistance due to the presence of the condenser. Since the current  $I$  is the quantity of electricity flowing through a cross-section of the conductor in unit time, the charge on the condenser is given by the time integral  $\int_0^t I dt$ , if we start to count time at the instant when the condenser has zero charge. Hence the potential difference of the two sides of the condenser which, neglecting the resistance of the connexions, is equal to the electromotive force of the generator, has the value

$$V = \frac{1}{C} \int_0^t I dt, \quad \text{so that} \quad \frac{dV}{dt} = \frac{I}{C}. \quad . \quad . \quad . \quad (18)$$

For a purely periodic applied potential we have, since  $dV/dt = i\omega V$ ,

$$V = \frac{I}{i\omega C}. \quad . \quad . \quad . \quad . \quad . \quad (19)$$

Hence the alternating current resistance offered by a capacity  $C$  is

given by  $1/i\omega C$ . In this case, since  $i$  in the denominator is equivalent to  $-i$  in the numerator, the applied electromotive force *lags behind* the current by a phase angle  $\pi/2$ .

By using these virtual resistances of self-induction and capacity we can treat any alternating current circuit like a direct current circuit. After completing the numerical computations, it is only necessary to draw the index diagram for current and electromotive force; this gives the amplitude and phase relationships between these quantities.

As an example, let us take a circuit containing ohmic resistance, self-induction and capacity connected to an alternator. By Kirchhoff's first law (p. 294),

$$V = V_R + V_L + V_C = I \left( R + i\omega L + \frac{1}{i\omega C} \right). \quad (20)$$

If we take  $I$  along the positive real axis in the index diagram,  $V_L$  is along the positive imaginary axis and  $V_C$  along the negative imaginary axis. According to fig. 1, the resultant electromotive force is of magnitude

$$V = I \sqrt{R^2 + \left( \omega L - \frac{1}{\omega C} \right)^2}. \quad (21)$$

This electromotive force leads the current by an angle  $\phi$ , so that we have  $V = V_0 e^{i\omega t}$ , and  $I = I_0 e^{i(\omega t - \phi)}$ , where

$$\tan \phi = \frac{\omega L - \frac{1}{\omega C}}{R}. \quad (22)$$

For an applied electromotive force of given amplitude  $V_0$ , the amplitude of the current is

$$I_0 = \frac{V_0}{\sqrt{R^2 + \left( \omega L - \frac{1}{\omega C} \right)^2}}. \quad (23)$$

Taking only the amplitude into consideration, the system has an apparent resistance of magnitude

$$\sqrt{R^2 + \left( \omega L - \frac{1}{\omega C} \right)^2}.$$

This quantity is called the *impedance* of the circuit. It has a minimum value when

$$\omega L = \frac{1}{\omega C} \quad \text{or} \quad \omega^2 = \frac{1}{LC}. \quad (24)$$

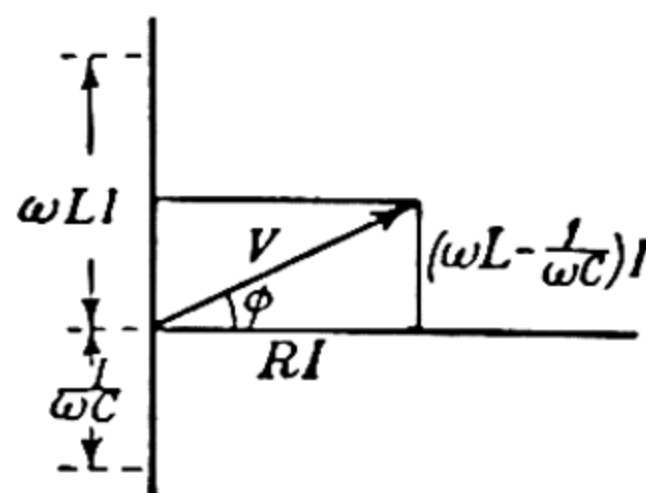


Fig. 1

In this case there is no difference of phase between  $V$  and  $I$ . This condition corresponds to *resonance*, and we shall find, in the next section, that the frequency of the applied potential in this case is the same as the natural frequency of the circuit.

In order to calculate the power, i.e. the amount of electrical work done per second, it is convenient to turn to the real part of the equations and write

$$V = V_0 \cos \omega t, \quad I = I_0 \cos(\omega t - \phi). \quad . \quad . \quad . \quad (25)$$

The power  $W_t$  at a given instant  $t$  is

$$W_t = VI = V_0 I_0 \cos \omega t \cos(\omega t - \phi). \quad . \quad . \quad . \quad (26)$$

It is more important to know the time average of the power,  $\bar{W}$ . We need only calculate the average over one period. Thus

$$\begin{aligned} \overline{\cos \omega t \cos(\omega t - \phi)} &= \frac{\omega}{2\pi} \int_0^{2\pi/\omega} (\cos \omega t \cos \omega t \cos \phi + \cos \omega t \sin \omega t \sin \phi) dt \\ &= \frac{1}{2} \cos \phi. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (27) \end{aligned}$$

Hence the average power is

$$\bar{W} = \frac{1}{2} V_0 I_0 \cos \phi. \quad . \quad . \quad . \quad . \quad . \quad . \quad (28)$$

The expression  $\cos \phi$  is known as the *power factor* of the circuit. If the impedance is purely inductive and capacitive ( $R=0$ ),  $\phi = \pm \pi/2$ , and no power is consumed in the circuit. In this case the current is said to be *wattless*. This shows why it is better to control alternating currents by choke coils (inductances) rather than by ohmic resistances.

The analogy between equation (28) and the expression for mechanical work (force into distance into cosine of the included angle), which becomes especially apparent with the use of the vector diagram, has no deeper significance and is of mnemonic value only.

*Ex. 91.* Calculate the impedance of an inductance and a capacity connected in parallel.

*Ex. 92.* Show that an oscillating circuit containing an iron-cored coil as self-inductance is a non-harmonic system to which the results of p. 101 may be applied.

#### 4. Non-stationary States (Transient Phenomena) in Alternating Current Circuits.

The relationships between current and electromotive force in alternating current circuits which we have derived by means of the vector diagram refer to the steady state which sets in after some time. The procedure of § 3 is not applicable to the transient conditions immediately following the completion of the circuit, for these phenomena depend upon the initial conditions. Since the results obtained



above place no constants of integration at our disposal, we have no way of satisfying the initial conditions. Mathematically, the solutions corresponding to the steady state are particular integrals of the differential equation, and we have now to find the general solution. We shall treat four different systems, beginning with the simplest and proceeding to the more complicated.

(a) *Free oscillations of a circuit containing only inductance and capacity*

Assume that the two plates of a charged condenser are joined by a conductor of negligible resistance. We must insert the expressions for the potentials due to self-induction and capacity (§ 3), in equation (20), p. 318, in the form of derivatives and integrals:

$$L \frac{dI}{dt} + \frac{1}{C} \int I dt = 0. \quad . . . . . (29)$$

Differentiation yields the familiar equation of simple harmonic vibrations

$$L \frac{d^2 I}{dt^2} + \frac{1}{C} I = 0, \quad . . . . . (30)$$

whose integral is  $I = ae^{i\omega_0 t} + be^{-i\omega_0 t} \quad . . . . . (31)$

where  $\omega_0 = \frac{1}{\sqrt{LC}}. \quad . . . . . (32)$

Using real functions, (31) may be written

$$I = A \cos \omega_0 t + B \sin \omega_0 t \quad \text{or} \quad I = I_0 \cos(\omega_0 t - \phi). \quad (31')$$

(b) *Free oscillations of a circuit containing inductance, capacity and resistance. Damping*

Here we have a term  $RI$  to be added to the left member of (29). Differentiating as above, we find

$$L \frac{d^2 I}{dt^2} + R \frac{dI}{dt} + \frac{1}{C} I = 0. \quad . . . . . (33)$$

This equation, too, is known from mechanics [Chap. V, § 8 (p. 95)]. The real form of the general solution obtained there is, for

$$\frac{1}{LC} > \frac{R^2}{4L^2}.$$

$$I = Ae^{-Rt/2L} \cos(\omega_0' t - \phi). \quad . . . . . (34)$$

The amplitude of the vibration decreases exponentially as the time increases, and the vibration is therefore said to be *damped*. The frequency of the characteristic oscillation of the damped system,  $\omega_0'$ , is smaller than that of the undamped circuit:

$$\omega_0' = \sqrt{\frac{1}{LC} - \frac{R^2}{4L^2}} < \omega_0 = \frac{1}{\sqrt{LC}}. \quad . \quad . \quad . \quad (35)$$

The ratio of two successive amplitudes, corresponding to one period, is constant, and is called the decrement. One usually deals with the natural logarithm of this quantity—the logarithmic decrement  $\delta$  (cf. p. 96):

$$\delta = \log \frac{e^{-\frac{R}{2L}t}}{e^{-\frac{R}{2L}(t+T)}} = \frac{R}{2L} T = \frac{R}{2L} \frac{2\pi}{\omega_0'}, \quad . \quad . \quad (36)$$

where  $T$  is the period. If the damping is not too great we may here use the frequency of the undamped vibration in place of the actual frequency of the damped system, in which case the logarithmic decrement becomes

$$\delta = \pi R \sqrt{\frac{C}{L}}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (37)$$

On the other hand, if  $\frac{1}{LC} < \frac{R^2}{4L^2}$

there is no imaginary part, and the solution is a purely exponential decreasing function

$$I = Ae^{-(\frac{R}{2L} + \sqrt{\frac{R^2}{4L^2} - \frac{1}{LC}})t} + Be^{-(\frac{R}{2L} - \sqrt{\frac{R^2}{4L^2} - \frac{1}{LC}})t}, \quad . \quad (38)$$

an aperiodic system. The limiting case  $1/LC = R^2/4L^2$  is also aperiodic; it corresponds to the mechanical situation treated on p. 95.

(c) *Damped circuit with periodic applied electromotive force*

The differential equation is

$$L \frac{dI}{dt} + RI + \frac{1}{C} \int I dt = V_0 e^{i\omega t}. \quad . \quad . \quad . \quad (39)$$

Differentiating with respect to the time,

$$L \frac{d^2 I}{dt^2} + R \frac{dI}{dt} + \frac{1}{C} I = i\omega V_0 e^{i\omega t}. \quad . \quad . \quad . \quad (40)$$

According to the theorem on p. 92, the general solution of this non-homogeneous equation is given by the sum of the general solution of the corresponding homogeneous equation, i.e. equation (33) above,

and a particular integral of the non-homogeneous equation. But we already know a particular integral, viz. that corresponding to the steady state, which was derived in § 3 (p. 318). For

$$\frac{1}{LC} > \frac{R^2}{4L^2},$$

the complete general solution is

$$I = e^{-\frac{R}{2L}t} (A \cos \omega_0' t + B \sin \omega_0' t) + \frac{V_0 \cos(\omega t - \phi)}{\sqrt{R^2 + \left(L\omega - \frac{1}{\omega C}\right)^2}}, \quad (41)$$

where

$$\tan \phi = \frac{L\omega - \frac{1}{\omega C}}{R}.$$

The part of this equation containing the exponential factor represents an oscillation having the angular frequency  $\omega_0'$  of the characteristic oscillation of the free damped circuit; the other term is an undamped periodic function having the angular frequency  $\omega$  of the applied electromotive force. At the beginning, the resultant vibration is a complicated oscillation which, however, soon attains the steady state given by the last term, the transient part being negligible after a short time.

(d) *Inductively coupled free circuits*

Of the many possible arrangements of coupled circuits which are of technical importance we select the simplest case for treatment here. Let us consider two circuits which are inductively coupled. Let the self-inductances be  $L_{11}$  and  $L_{22}$  respectively, and let the mutual inductance be  $L_{12}$ . Here, again, we shall carry the calculations only far enough to reveal the new principles involved in this type of coupled system.

The differential equation for the first circuit becomes after differentiating once with respect to the time:

$$L_{11} \frac{d^2 I_1}{dt^2} + L_{12} \frac{d^2 I_2}{dt^2} + \frac{I_1}{C_1} = 0. \quad . \quad . \quad . \quad (42)$$

That of the second circuit becomes

$$L_{12} \frac{d^2 I_1}{dt^2} + L_{22} \frac{d^2 I_2}{dt^2} + \frac{I_2}{C_2} = 0. \quad . \quad . \quad . \quad (43)$$

In order to eliminate  $I_2$ , we differentiate equation (43) twice with respect to  $t$ , obtaining

$$L_{12} \frac{d^4 I_1}{dt^4} + L_{22} \frac{d^4 I_2}{dt^4} + \frac{1}{C_2} \frac{d^2 I_2}{dt^2} = 0. \quad . \quad . \quad . \quad (44)$$



But from (42), 
$$\frac{d^2 I_2}{dt^2} = -\frac{1}{L_{12}} \left( L_{11} \frac{d^2 I_1}{dt^2} + \frac{1}{C_1} I_1 \right), \quad . . . \quad (45)$$

and after differentiating twice,

$$\frac{d^4 I_2}{dt^4} = -\frac{1}{L_{12}} \left( L_{11} \frac{d^4 I_1}{dt^4} + \frac{1}{C_1} \frac{d^2 I_1}{dt^2} \right). \quad . . . \quad (46)$$

Substituting these values in equation (44), and using the abbreviations

$$a = L_{11} L_{22} - L_{12}^2, \quad b = \frac{L_{11}}{C_2} + \frac{L_{22}}{C_1}, \quad c = \frac{1}{C_1 C_2}, \quad . \quad (47)$$

we obtain the fourth-order differential equation

$$a \frac{d^4 I_1}{dt^4} + b \frac{d^2 I_1}{dt^2} + c I_1 = 0. \quad . . . \quad (48)$$

We assume a solution of the form

$$I_1 = A e^{\lambda t} \quad . . . \quad (49)$$

and obtain the characteristic equation

$$a \lambda^4 + b \lambda^2 + c = 0. \quad . . . \quad (50)$$

Of the four roots, two are essentially different:

$$\lambda_1 = i \sqrt{\frac{b}{2a} + \sqrt{\frac{b^2}{4a^2} - \frac{c}{a}}} = i \omega_1, \quad . . . \quad (51)$$

$$\lambda_2 = i \sqrt{\frac{b}{2a} - \sqrt{\frac{b^2}{4a^2} - \frac{c}{a}}} = i \omega_2. \quad . . . \quad (52)$$

The other two roots differ from these only in sign. According to *Ex. 90* (p. 316),  $a$  is always positive, and  $b$  and  $c$  are essentially positive, while  $b^2 - 4ac$  is a sum of squares. Hence, since  $L_{12} \neq 0$ , we always have

$$0 < \frac{b^2}{4a^2} - \frac{c}{a} < \frac{b^2}{4a^2}; \quad . . . \quad (53)$$

That is, the two frequencies  $\omega_1$  and  $\omega_2$  are always real. By eliminating  $d^2 I_2/dt^2$  between (42) and (43) we see that there exist two simultaneous frequencies  $\omega_1$  and  $\omega_2$  for  $I_2$  as well as for  $I_1$ . Neither of these frequencies coincides with the natural frequencies of the separate circuits. The solution for  $I_1$  thus has the form

$$I_1 = A e^{i \omega_1 t} + B e^{-i \omega_1 t} + C e^{i \omega_2 t} + D e^{-i \omega_2 t}. \quad . . . \quad (54)$$

The co-existence of two vibrations of different frequency implies the occurrence of *beats*, i.e. the appearance of slow variations of amplitude

as may be seen at once from the following trigonometric identity:

$$\cos \omega_1 t + \cos \omega_2 t = 2 \cos \left( \frac{\omega_1 - \omega_2}{2} t \right) \cos \left( \frac{\omega_1 + \omega_2}{2} t \right).$$

The right member represents a vibration of angular frequency  $(\omega_1 + \omega_2)/2$  whose amplitude is "modulated" (cf. p. 57) with an angular frequency  $(\omega_1 - \omega_2)/2$ . Since the circuits are assumed to be undamped, the energy of circuit 1 cannot be destroyed when its amplitude is reduced to zero, but must be transmitted to the other circuit. The beats occurring in the two circuits must therefore have a phase difference of  $\pi/2$ , so that minimum amplitude in the first circuit corresponds to maximum amplitude in the second, and the sum of the energies (squares of the amplitudes) remains constant. This is actually confirmed by the complete calculation, which will not be gone into here.

### 5. Resistance and Inductance of Wires for Alternating Currents. Skin Effect.

All low-frequency electromagnetic phenomena in homogeneous media are determined by the Law of Induction (p. 312), together with equation (9) (p. 301), which was derived for steady currents, and expresses the connexion between current and magnetic field. If several different media are present, the boundary conditions of the continuity of the tangential components of  $\mathbf{E}$  and  $\mathbf{H}$  must also be taken into account. In view of the meaning of the curl, the above two equations really represent a system of six first-order partial differential equations for the six components of  $\mathbf{E}$  and  $\mathbf{H}$ . Of the enormous number of solutions of such a system, we must select that which corresponds to the problem under consideration. This selection, and not the calculation of the integrals, is the difficult matter. We consider the following very simple case. Let there be given a very long, straight wire of finite cross-section to whose ends an alternating difference of potential is applied. For direct current, the current density would be the same at all points within the wire, provided the conductor is made of homogeneous material. For alternating current, the current density is greater near the surface of the wire, on account of inductive effects. When the frequency is very great, the current is almost entirely confined to the outer layers. This phenomenon is known as the *skin effect*.

The dependence of the current density upon the distance  $\rho$  from the axis may be calculated with little difficulty from the two differential equations:

$$\text{curl } \mathbf{H} = \frac{4\pi \mathbf{i}}{c}; \quad \text{curl } \mathbf{E} = \frac{1}{\sigma} \text{curl } \dot{\mathbf{i}} = -\frac{\mu}{c} \frac{\partial \mathbf{H}}{\partial t}. \quad \cdot \cdot \cdot (55)$$

**H** may be eliminated as follows. Take the curl of the second equation and differentiate the first with respect to  $t$ . Since we are dealing with bodies at rest, the derivatives with respect to the co-ordinates (in taking the curl) and those with respect to the time are entirely independent. Hence

$$\frac{\partial}{\partial t} \text{curl } \mathbf{H} = \text{curl } \frac{\partial \mathbf{H}}{\partial t}.$$

If we eliminate this term from the two equations obtained, we find

$$\text{curl curl } \mathbf{i} = \text{grad div } \mathbf{i} - \Delta \mathbf{i} = - \frac{4\pi\sigma\mu}{c^2} \frac{\partial \mathbf{i}}{\partial t}.$$

Since  $\text{div } \mathbf{i} = 0$ , this equation reduces to

$$\Delta \mathbf{i} = \frac{4\pi\sigma\mu}{c^2} \frac{\partial \mathbf{i}}{\partial t}. \quad . \quad . \quad . \quad . \quad . \quad (56)$$

On account of symmetry, the vector  $\mathbf{i}$  in the infinitely long wire has the direction of the axis of the wire. We take this to be the  $z$ -axis. Then we may write

$$\mathbf{i} = u(\rho, t)\mathbf{k}. \quad . \quad . \quad . \quad . \quad . \quad (57)$$

We thus obtain a scalar equation which has the following form in cylindrical co-ordinates:

$$\frac{\partial^2 u}{\partial \rho^2} + \frac{1}{\rho} \frac{\partial u}{\partial \rho} = \frac{4\pi\sigma\mu}{c^2} \frac{\partial u}{\partial t}. \quad . \quad . \quad . \quad . \quad (58)$$

We are looking for a solution  $u$  which is a periodic function of the time, and so we put

$$u = f(\rho)e^{i\omega t} \quad . \quad . \quad . \quad . \quad . \quad (59)$$

and obtain the following ordinary differential equation for  $f$ :

$$\frac{d^2 f}{d\rho^2} + \frac{1}{\rho} \frac{df}{d\rho} - i \frac{4\pi\sigma\mu\omega}{c^2} f = 0. \quad . \quad . \quad . \quad . \quad (60)$$

In this equation we set

$$- \frac{4\pi\sigma\mu\omega i}{c^2} = a^2, \quad . \quad . \quad . \quad . \quad . \quad (61)$$

and introduce a new variable  $x = a\rho$ . Then

$$\frac{df}{d\rho} = \frac{df}{dx} a, \text{ \&c.} \quad . \quad . \quad . \quad . \quad . \quad (62)$$

and we obtain

$$\frac{d^2 f}{dx^2} + \frac{1}{x} \frac{df}{dx} + f = 0. \quad . \quad . \quad . \quad . \quad (63)$$

But this is the differential equation for the Bessel functions of order



zero.\* The only one of these functions of use here—the one continuous at the origin—is denoted by  $J_0(x)$ . Hence

$$u = AJ_0(a\rho)e^{i\omega t} = AJ_0\left(\sqrt{-i}\sqrt{4\pi\sigma\mu\omega}\frac{\rho}{c}\right)e^{i\omega t}, \quad (64)$$

where  $A$  is the amplitude factor.

The real part and the imaginary part of the function  $J_0(x\sqrt{-i})$  occur in various problems and are tabulated as new functions of  $x$ . They are known as “ber ( $x$ )” and “bei ( $x$ )” respectively:

$$\left. \begin{aligned} (Re)[J_0 x \sqrt{-i}] &= \text{ber } x \\ (Im)[J_0 x \sqrt{-i}] &= \text{bei } x \end{aligned} \right\} \quad \cdot \cdot \cdot \cdot \quad (65)$$

Using these functions, our solution may be written

$$u = A \left[ \text{ber} \left( \sqrt{4\pi\sigma\mu\omega} \frac{\rho}{c} \right) + i \text{bei} \left( \sqrt{4\pi\sigma\mu\omega} \frac{\rho}{c} \right) \right] e^{i\omega t} \quad (66)$$

or 
$$u = A \sqrt{\text{ber}^2 \left( \sqrt{4\pi\sigma\mu\omega} \frac{\rho}{c} \right) + \text{bei}^2 \left( \sqrt{4\pi\sigma\mu\omega} \frac{\rho}{c} \right)} e^{i(\omega t + \phi)}, \quad (66')$$

where 
$$\tan \phi = \frac{\text{bei} \left( \sqrt{4\pi\sigma\mu\omega} \frac{\rho}{c} \right)}{\text{ber} \left( \sqrt{4\pi\sigma\mu\omega} \frac{\rho}{c} \right)}.$$

That is, not only the amplitude, but also the phase depends upon the radius. For large values of  $\mu$  (iron wire) both effects can be shown even for currents of very low frequency (1 cycle/sec.). One can obtain a picture of the process by drawing the vector for the current density corresponding to various values of  $\rho$  (fig. 2). In

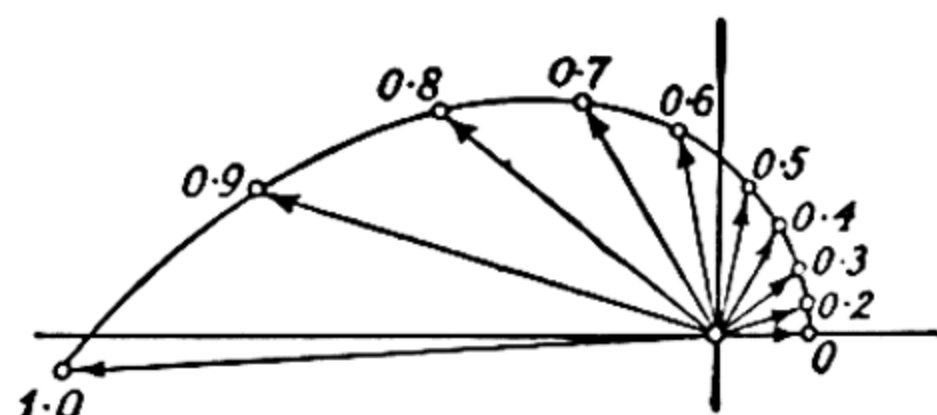


Fig. 2

order to determine the amplitude factor  $A$  we have to integrate  $i$  over the cross-section, which gives the total current as a function of  $A$ , so that  $A$  in turn may be evaluated in terms of  $I$ .

\* See p. 810 et seq.

## CHAPTER XVII

### RAPIDLY ALTERNATING ELECTROMAGNETIC FIELDS: I. PROPAGATION IN HOMOGENEOUS ISOTROPIC MEDIA

#### 1. The Electrical Analogue of the Law of Induction.

The Law of Induction states that, in a vacuum, any change of the magnetic flux through a surface bounded by a curve results in the formation of an electrical "vortex", i.e. a finite value of the line integral  $\oint \mathbf{E} d\mathbf{s}$ . The reciprocal relationships between electrical and magnetic phenomena already encountered many times would lead us to expect that a change in electrical flux similarly would cause a magnetic "vortex". Thus far we have encountered magnetic curl only when conduction currents were present. We write as a tentative form of relationship, for a vacuum,

$$\text{curl } \mathbf{H} = \frac{1}{c} \frac{\partial \mathbf{E}}{\partial t}, \text{ or (M.K.S.) } K_0 \frac{\partial \mathbf{E}}{\partial t}. \quad \cdot \cdot \cdot (1)$$

This equation was derived by Maxwell from the notion that the volume elements of the "ether" are polarized when a field is applied, just as are those of a material dielectric (p. 279). It may be seen readily that electric charges are displaced when a volume element is polarized, and hence that a current flows while the polarization is being brought about.

Imagine a small cylinder of unit cross-section and altitude  $dh$  in which a polarization  $\mathbf{P}$  is generated parallel to the axis. The apparent charges on the transverse surfaces of the cylinder amount to  $\sigma_+ = +|\mathbf{P}|$  on the upper side and  $\sigma_- = -|\mathbf{P}|$  on the lower, according to p. 282. If  $\mathbf{P}$  changes by  $d\mathbf{P}$ , the density of charge is altered by  $+|d\mathbf{P}|$  and  $-|d\mathbf{P}|$  respectively. If this charge is to reach the bounding surface it must pass through the cross-section immediately below or above. Hence a positive current  $d\mathbf{P}/dt$  flows in the upper half and a negative current  $-d\mathbf{P}/dt$  flows in the lower half, but in the opposite direction, so that a "displacement current" of

$$i_p = \frac{1}{c} \frac{d\mathbf{P}}{dt}$$

electromagnetic units flows during the time the polarization is changing. But Maxwell's concept of a polarizable ether built up of electric charges is no longer tenable, and we are justified in assuming the existence of a true displacement current only in the case of material dielectrics. Hence we have thus far no rigorous basis for the term

$$\frac{1}{c} \frac{\partial \mathbf{E}}{\partial t},$$

especially with regard to the sign. We shall therefore justify it for the present only by noting that all its consequences are confirmed by experience. Later on we shall see how the field equations may be deduced by a method which is based on very few hypotheses. Hence in a conducting dielectric medium we have three contributions to the current producing a magnetic field, so that we may write in general

$$\text{curl } \mathbf{H} = \frac{4\pi i}{c} + \frac{1}{c} \frac{\partial \mathbf{E}}{\partial t} + \frac{4\pi}{c} \frac{\partial \mathbf{P}}{\partial t} = \frac{4\pi i}{c} + \frac{1}{c} \frac{\partial \mathbf{E}}{\partial t} + \frac{4\pi}{c} \kappa \frac{\partial \mathbf{E}}{\partial t}. \quad (2)$$

The last two terms are usually combined into  $\frac{K}{c} \frac{\partial \mathbf{E}}{\partial t}$ , so that we write

$$\text{curl } \mathbf{H} = \frac{K}{c} \frac{\partial \mathbf{E}}{\partial t} + \frac{4\pi}{c} \sigma \mathbf{E}, \quad . . . . . (2a)$$

$$\text{or, in M.K.S. units,} \quad \text{curl } \mathbf{H} = KK_0 \frac{\partial \mathbf{E}}{\partial t} + \sigma \mathbf{E}. \quad . . . . . (2b)$$

Following Maxwell's ideas, we refer to the term

$$\frac{K}{c} \frac{\partial \mathbf{E}}{\partial t}$$

as the *displacement current*, to distinguish it from the actual conduction current. The above equation (2a), together with equation (3a) (p. 312) [or (2b) and (3b)], constitute Maxwell's equations for the electromagnetic field.

## 2. The Wave Equation for the Propagation of Fields in Dielectrics.

For isotropic dielectrics at rest the following equations hold:

$$\text{curl } \mathbf{H} = \frac{K}{c} \frac{\partial \mathbf{E}}{\partial t} \text{ or } KK_0 \frac{\partial \mathbf{E}}{\partial t}. \quad . . . . . (I)$$

$$\text{curl } \mathbf{E} = -\frac{\mu}{c} \frac{\partial \mathbf{H}}{\partial t} \text{ or } -KK_0 \frac{\partial \mathbf{H}}{\partial t}. \quad . . . . . (II)$$

If, as we shall assume, no volume charges are present, we have in



addition the equations of continuity of the electrical and magnetic fields, viz.

$$\operatorname{div} \mathbf{E} = 0 \quad \text{and} \quad \operatorname{div} \mathbf{H} = 0. \quad . \quad . \quad . \quad . \quad (III)$$

Either  $\mathbf{E}$  or  $\mathbf{H}$  may be eliminated from the Maxwell equations by the method of p. 325. If we choose to eliminate  $\mathbf{H}$ , we take the curl of the second equation, while we differentiate the first with respect to  $t$ . Using (III) we then obtain

$$\Delta \mathbf{E} = \frac{K\mu}{c^2} \frac{\partial^2 \mathbf{E}}{\partial t^2}, \quad \text{or} \quad \Delta \mathbf{E} = KK_0\mu\mu_0 \frac{\partial^2 \mathbf{E}}{\partial t^2}. \quad . \quad . \quad . \quad (3)$$

This is the differential equation of a wave-like propagation of the components of  $\mathbf{E}$ . By eliminating  $\mathbf{E}$ , we obtain an exactly similar equation for  $\mathbf{H}$ . Of the tremendous manifold of phenomena described by this differential equation we now consider the simplest case—that of a plane wave. We understand this to mean a wave for which all quantities depend only upon one co-ordinate, say  $x$ . The components of  $\mathbf{E}$  have constant values in planes normal to the  $x$ -axis. Mathematically, this means the vanishing of all derivatives  $\partial/\partial y$  and  $\partial/\partial z$ . Hence from (III),

$$\operatorname{div} \mathbf{E} = \frac{\partial E_x}{\partial x} + \frac{\partial E_y}{\partial y} + \frac{\partial E_z}{\partial z} = 0,$$

it follows that

$$\frac{\partial E_x}{\partial x} = 0. \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

Hence the electric field can have no component in the direction of the  $x$ -axis. Since we are dealing with a wave motion, a superimposed field which has the same value at all points does not interest us here; we may conclude that equation (4) implies that  $E_x = 0$ . In the same way we find that  $H_x = 0$ . Hence *the vanishing of the divergence of  $\mathbf{E}$  and  $\mathbf{H}$  implies the transversality of the plane wave.*

We obtain the equations

$$\frac{\partial^2 E_y}{\partial x^2} = \frac{K\mu}{c^2} \frac{\partial^2 E_y}{\partial t^2} \quad \text{and} \quad \frac{\partial^2 E_z}{\partial x^2} = \frac{K\mu}{c^2} \frac{\partial^2 E_z}{\partial t^2}. \quad . \quad . \quad (5)$$

for the components  $E_y$  and  $E_z$ . We assume a simple periodic dependence on the time:

$$E_y = f(x)e^{i\omega t} \quad \text{and} \quad E_z = g(x)e^{i(\omega t - \delta)}, \quad . \quad . \quad . \quad (6)$$

where the phase constant  $\delta$  is included to ensure generality. Then, by (5)

$$\frac{d^2 f}{dx^2} + \frac{\omega^2 K\mu}{c^2} f = 0 \quad \text{and} \quad \frac{d^2 g}{dx^2} + \frac{\omega^2 K\mu}{c^2} g = 0. \quad . \quad . \quad (7)$$

The integrals of these well-known equations of simple harmonic vibrations are

$$f = Ae^{\pm i\omega \sqrt{K\mu} x/c} \quad \text{and} \quad g = Be^{\pm i\omega \sqrt{K\mu} x/c}. \quad (8)$$

Using the negative sign, we then have

$$E_y = Ae^{i\omega(t - \sqrt{K\mu} x/c)} \quad \text{and} \quad E_z = Be^{i\omega(t - \sqrt{K\mu} x/c) - i\delta}. \quad (9)$$

This is the equation of a plane wave propagated in the direction of the positive  $x$ -axis (the negative sign would signify the opposite direction) with a wave velocity

$$v = c/\sqrt{K\mu}, \quad \text{or} \quad v = 1/\sqrt{KK_0\mu\mu_0}. \quad (10)$$

In the C.G.S. system  $v = c$  for a vacuum, so that the factor  $c$  introduced on p. 300 has a highly significant physical meaning. It represents the speed of electromagnetic waves in a vacuum, as anticipated there. In the M.K.S. system  $c = 1/\sqrt{K_0\mu_0}$ , as is evident from the dimensions of  $K_0\mu_0$ .

$E_y$  and  $E_z$  are independent of each other, so that the phase difference is arbitrary. The most general form is thus the elliptically polarized wave in which the end point of the vector  $\mathbf{E}$  describes an ellipse at any given place; if this motion is confined to a straight line we have a plane polarized wave.

In order to calculate the components of  $\mathbf{H}$  we are not permitted to start with the wave equation for this quantity; this would furnish us with an excess of constants. We must observe that  $\mathbf{E}$  and  $\mathbf{H}$  are connected by the field equations. For  $E_z = 0$ , the second one gives

$$\frac{\partial H_y}{\partial t} = 0, \quad \frac{\mu}{c} \frac{\partial H_z}{\partial t} = -\frac{\partial E_y}{\partial x} = \frac{\sqrt{K\mu}}{c} i\omega A e^{i\omega(t - \sqrt{K\mu} x/c)}. \quad (11)$$

Integrating,

$$H_y = 0, \quad H_z = \sqrt{\frac{K}{\mu}} A e^{i\omega(t - \sqrt{K\mu} x/c)} = \sqrt{\frac{K}{\mu}} E_y. \quad (12)$$

Since we are no more interested in a temporally constant field than in a spatially constant one, we may set the integration constants of (12), which are independent of time (functions of  $x$ ) equal to zero. Thus we have

$$E_y H_y + E_z H_z = 0,$$

so that  $\mathbf{E}$  is perpendicular to  $\mathbf{H}$  in a plane polarized wave. Let us designate the direction of propagation by the unit vector  $\mathbf{n}$ . Then  $\mathbf{E}$ ,  $\mathbf{H}$  and  $\mathbf{n}$  form a right-handed orthogonal triad, in this order.

Electromagnetic waves embrace an enormous range of physical phenomena, the essential unity of which remained unrecognized for many years, chiefly because the methods of production and the effects of the waves differ widely according to the value of the frequency  $\nu$  (or the vacuum wave-length  $\lambda = c/\nu$ ), as the accompanying short

table (fig. 1) shows. The entire range, with certain overlappings of the methods of generation and of investigation, is now known without a break.

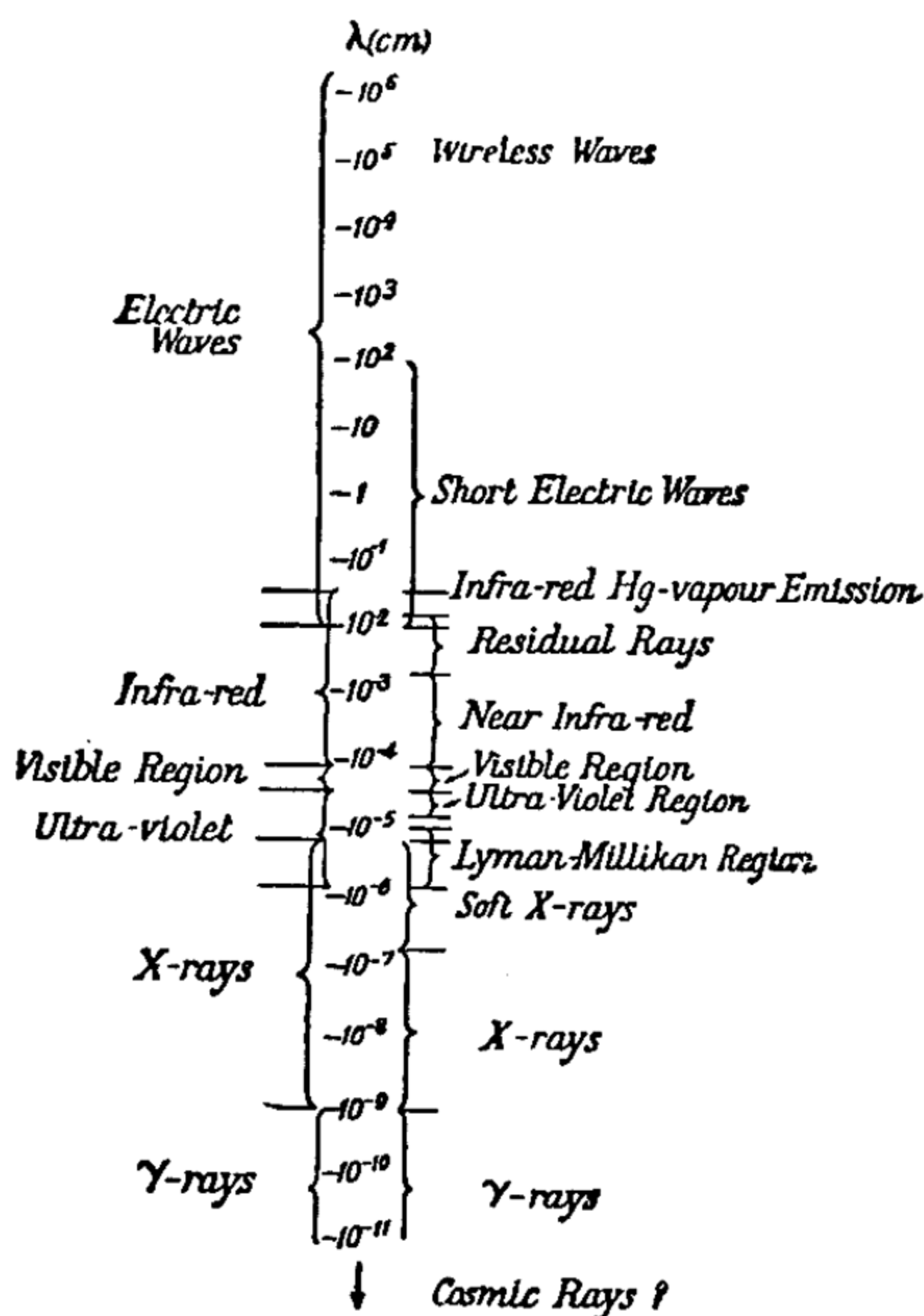


Fig. 1

The discrepancy is explained by the fact that the polarization of a medium (which determines the value of  $K$ ) becomes dependent upon the frequency for rapidly alternating fields. Down to wave-lengths of about 1 cm. the static value of the dielectric constant may be used for all substances. For shorter waves (higher frequencies) dependence on frequency comes in. This will be explained in the second part of the Theory of Electricity (p. 452).

### 3. The Poynting Vector of Energy Flow.

The energy density in the electromagnetic field has the value

$$u = \frac{1}{8\pi} (K\mathbf{E}^2 + \mu\mathbf{H}^2), \text{ or } u = \frac{1}{2} (KK_0\mathbf{E}^2 + \mu\mu_0\mathbf{H}^2). \quad (14)$$

Since by (12),  $|\mathbf{H}| = \sqrt{\frac{K}{\mu}} |\mathbf{E}|$  or  $= \sqrt{\frac{KK_0}{\mu\mu_0}} |\mathbf{E}|$

In optics, the ratio of the phase velocity  $c$  in vacuum to that in any other medium is known as the *index of refraction*  $n$  of the medium. Thus, by equation (10),

$$K\mu = n^2. \quad (13)$$

This equation is known as Maxwell's relation. But, as even the familiar example of water ( $K = 80$ ,  $\mu \approx 1$ ,  $n = 1.33$ ) shows, this relation is not always satisfied if we take  $n$  to be the optical index of refraction. If (13) were satisfied down to arbitrarily short waves there could be no dispersion at all, i.e. no variation of  $n$  with wave-length, if  $K$  and  $\mu$  are not dependent upon frequency.



for a plane wave,  $K\mathbf{E}^2 = \mu\mathbf{H}^2$ , or  $KK_0\mathbf{E}^2 = \mu\mu_0\mathbf{H}^2$ . . . . (15)

Thus the energy is half electrical and half magnetic in this case. Now we know that energy is transported by every electromagnetic wave. In order to verify this experimentally we need only convert the energy into heat at a given point and determine calorimetrically the rate at which energy is being delivered. If the wave advances with a velocity  $v$ , the amount of energy streaming through unit area of a surface normal to the direction of propagation in the time  $dt$  is that residing in the space within a rectangular parallelepiped of  $1 \text{ cm.}^2$  base and altitude  $vdt$ , i.e.

$$dU = \frac{vdt}{8\pi} (K\mathbf{E}^2 + \mu\mathbf{H}^2) = \frac{cdt}{\sqrt{K\mu}} \cdot \frac{1}{4\pi} K\mathbf{E}^2 = \frac{cdt}{4\pi} |\mathbf{E}| \cdot |\mathbf{H}|. \quad (16)$$

We may therefore represent the energy flow by a vector in the direction of propagation  $\mathbf{n}$  whose magnitude is equal to the energy flow per unit time. This is the *Poynting Vector*

$$\mathbf{S} = \frac{c}{4\pi} [\mathbf{E}\mathbf{H}], \text{ or } \mathbf{S}_{\text{M.K.S.}} = [\mathbf{E}\mathbf{H}], \quad . . . (17)$$

so named after its discoverer. Since  $\mathbf{E}$  and  $\mathbf{H}$  are periodic functions of the time, the vector  $\mathbf{S}$  at a given place will also experience periodic fluctuations between zero and a maximum value. A change of sign never occurs, however, since  $\mathbf{E}$  and  $\mathbf{H}$  are always in phase [cf. (12)]. The Poynting vector is always in the direction of propagation.

The Poynting vector  $\mathbf{S}$  has a much more general significance than might be expected from the way it was introduced. We never come into conflict with experience by assuming that this vector may be applied in all cases. It holds for all simultaneous electric and magnetic fields—even for steady fields. For example, consider a cylindrical wire bent into the form of a circle, the ends being joined to a source of direct electromotive force. Within the wire, the electric field is along the axis, while the magnetic lines of force are circles whose planes are normal to the wire. Hence the energy flux vector at the surface of the conductor is normal to the surface and, moreover, is directed inwards no matter which way the current flows. Thus there is a continuous flow of energy from the surrounding dielectric into the wire, and it is readily shown that the amount of energy entering unit length of wire each second is exactly equal to the Joule heat developed in that portion of the wire per unit time. The ultimate source of energy is the battery. From the atomic standpoint, however, this interpretation sounds artificial.

*Ex. 93.* The amount of solar energy received by the earth is about  $2 \text{ cal./min./cm.}^2$ . What are the amplitudes of the electric and magnetic fields of the radiation?

## 4. Propagation of Electromagnetic Waves in Conducting Media.

For media in which there exists a non-vanishing conductivity the field equations are

$$\text{curl } \mathbf{H} = \frac{K}{c} \frac{\partial \mathbf{E}}{\partial t} + \frac{4\pi\sigma}{c} \mathbf{E}. \quad \dots \quad (\text{I})$$

$$\text{curl } \mathbf{E} = -\frac{\mu}{c} \frac{\partial \mathbf{H}}{\partial t}. \quad \dots \quad (\text{II})$$

$$\text{div } \mathbf{E} = 0, \quad \text{div } \mathbf{H} = 0. \quad \dots \quad (\text{III})$$

If  $\mathbf{H}$  is again eliminated by the method of § 2 (p. 329), we have the following equation for  $\mathbf{E}$ :

$$\Delta \mathbf{E} = \frac{K\mu}{c^2} \frac{\partial^2 \mathbf{E}}{\partial t^2} + \frac{4\pi\sigma\mu}{c^2} \frac{\partial \mathbf{E}}{\partial t},$$

or 
$$\Delta \mathbf{E} = KK_0\mu\mu_0 \frac{\partial^2 \mathbf{E}}{\partial t^2} + \mu\mu_0 \frac{\partial \mathbf{E}}{\partial t} \sigma. \quad \dots \quad (18)$$

We again consider plane waves propagated in the direction of the positive  $x$ -axis, and set

$$\frac{\partial}{\partial y} = 0, \quad \frac{\partial}{\partial z} = 0,$$

from which it follows that  $E_x = H_x = 0$ . Assume again that  $E_y$  and  $E_z$  are simple periodic functions of the time:

$$E_y = f(x)e^{i\omega t} \quad \text{and} \quad E_z = g(x)e^{i\omega t}. \quad \dots \quad (19)$$

For simplicity of notation, we restrict our discussion to plane polarized waves whose electric vector is in the  $xy$ -plane; hence we put  $E_z = 0$ . By substitution in the wave equation we obtain the following differential equation for the point function  $f$ :

$$\frac{d^2 f}{dx^2} + \left( \frac{K\mu\omega^2}{c^2} - i \frac{4\pi\sigma\mu\omega}{c^2} \right) f = 0, \quad \dots \quad (20)$$

the solution of which is

$$f(x) = Ae^{\pm i\sqrt{K\mu\omega^2 - i4\pi\sigma\mu\omega}x/c}. \quad \dots \quad (21)$$

If  $\sigma = 0$ , we again have the case of an undamped wave. In this instance we must choose the negative sign for a wave propagated in the positive  $x$ -direction; hence the negative sign is to be used here also. We now introduce the following notation for convenience:

$$\sqrt{K\mu\omega^2 - i4\pi\sigma\mu\omega} = \omega n(1 - i\kappa). \quad \dots \quad (22)$$

Squaring and equating reals and imaginaries separately, we obtain equations for  $n$  and  $\kappa$ :

$$\left. \begin{aligned} n^2 - n^2 \kappa^2 &= K\mu \\ n^2 \kappa &= \frac{2\pi\mu\sigma}{\omega} = \frac{\mu\sigma}{\nu} = \mu\sigma T \end{aligned} \right\}, \quad \cdot \cdot \cdot \cdot \quad (23)$$

where  $T$  is the period of vibration. Solving simultaneously we obtain

$$\left. \begin{aligned} n &= \sqrt{\frac{\mu}{2} (\sqrt{K^2 + 4\sigma^2 T^2} + K)} \\ n\kappa &= \sqrt{\frac{\mu}{2} (\sqrt{K^2 + 4\sigma^2 T^2} - K)} \end{aligned} \right\} \cdot \cdot \cdot \cdot \quad (24)$$

Using the notation of (22) in (21), equation (19) then gives

$$E_y = A e^{i\omega t} e^{-i\omega n x/c} e^{-\omega \kappa x/c} = A e^{-\omega \kappa x/c} e^{i\omega(t - n x/c)}. \quad (25)$$

The imaginary term  $-i n \kappa$  is thus seen to furnish a real negative exponential function, i.e. a damping factor, while the velocity of propagation of the phase is determined by  $n$ . Thus the use of this same  $n$  for the index of refraction in insulating media is justified. The meaning of  $\kappa$  is as follows. Since

$$\frac{n\omega}{c} = \frac{2\pi\nu}{\lambda} = \frac{2\pi}{\lambda},$$

the damping factor becomes  $e^{-2\pi\kappa x/\lambda}$ .

For  $x = \lambda$ , the factor becomes

$$e^{-2\pi\kappa}.$$

That is, in travelling a distance equal to the wave-length in a vacuum, the amplitude diminishes by a factor

$$e^{2\pi\kappa}.$$

We now calculate the vector  $\mathbf{H}$ , using equation (II) (p. 333). As in the case of insulating media, only a  $z$ -component of  $\mathbf{H}$  remains. Using (25), this is given by

$$\frac{\partial H_z}{\partial t} = -\frac{c}{\mu} \frac{\partial E_y}{\partial x} = \frac{\omega}{\mu} (n\kappa + in) A e^{-\omega \kappa n x/c} e^{i\omega(t - n x/c)}, \quad (26)$$

$$\text{i.e.} \quad H_z = \frac{1}{\mu} n(1 - i\kappa) A e^{-\omega \kappa n x/c} e^{i\omega(t - n x/c)} = \frac{1}{\mu} n(1 - i\kappa) E_y. \quad (27)$$

Now from the relationships in the complex plane,

$$1 - i\kappa = \sqrt{1 + \kappa^2} e^{-i\gamma}, \text{ and } \tan \gamma = \kappa. \quad \cdot \cdot \cdot \cdot \quad (28)$$



Hence

$$H_z = \frac{n}{\mu} \sqrt{1 + \kappa^2} \cdot E_y e^{-i\gamma} = \frac{n}{\mu} \sqrt{1 + \kappa^2} A e^{-\omega \kappa n x/c} e^{i\omega(t - nx/c) - i\gamma}. \quad (29)$$

Thus we find that there is a difference in phase between the electric and magnetic vectors in conducting media:  $\mathbf{H}$  lags behind  $\mathbf{E}$  by the angle  $\gamma$ . According to the discussion in this section, every conductor should absorb electromagnetic waves, while all insulators should transmit them. However, we can give at once examples which contradict this: A solution of NaCl, in spite of high electrical conductivity, is perfectly transparent, while vulcanite, which is a good insulator, is opaque. The explanation is again, as in the case of the index of refraction, variation with frequency—here, variation of the conductivity. In calculations we may use the static value of the conductivity down to a wave-length of about 1 cm. for all substances. A possible solution of the difficulty will be discussed in the second part of the Theory of Electricity (§ 2, p. 451).

### 5. Hertz's Solution of the Field Equations. The Hertzian Oscillator.

Up to now we have considered plane waves without inquiring into their origin. The waves actually come from spatially limited sources, whether they are wireless waves emitted from an antenna which is small compared with the distances spanned by the signals, or light waves emitted by the radiating atoms or molecules of a gaseous source of light. For this reason, the waves are approximately plane only at great distances from the source. H. Hertz found a solution of the field equations which provides a unified representation of source and radiation field in the whole of space. If we confine our attention to a vacuum (free space), the appropriate field equations are

$$\text{curl } \mathbf{H} = \frac{1}{c} \frac{\partial \mathbf{E}}{\partial t}. \quad \cdot \cdot \cdot \cdot \cdot \cdot \quad (\text{I})$$

$$\text{curl } \mathbf{E} = -\frac{1}{c} \frac{\partial \mathbf{H}}{\partial t}. \quad \cdot \cdot \cdot \cdot \cdot \cdot \quad (\text{II})$$

$$\text{div } \mathbf{E} = 0, \quad \text{div } \mathbf{H} = 0. \quad \cdot \cdot \cdot \cdot \cdot \cdot \quad (\text{III})$$

In determining the magnetic fields accompanying steady currents, it proved expedient to express  $\mathbf{H}$  as the curl of an auxiliary vector  $\mathbf{V}$ , the vector potential, which bore a simple relation to the distribution of currents. We proceed here in a similar manner. The formulæ become appreciably simpler if we choose as vector potential the time derivative of an auxiliary vector  $\mathbf{Z}$  and supply also the factor  $1/c$ . We thus put

$$\mathbf{H} = \frac{1}{c} \text{curl } \frac{\partial \mathbf{Z}}{\partial t}. \quad \cdot \cdot \cdot \cdot \cdot \cdot \quad (30)$$

Since the order of the differentiations can be reversed, we may also write

$$\mathbf{H} = \frac{1}{c} \frac{\partial}{\partial t} (\text{curl } \mathbf{Z}). \quad . . . . . (30')$$

Substituting in (I), and integrating with respect to  $t$ ,

$$\mathbf{E} = \text{curl curl } \mathbf{Z} = \text{grad div } \mathbf{Z} - \Delta \mathbf{Z}. \quad . . . (31)$$

In order that (II) may be satisfied we must have

$$\text{curl } \mathbf{E} = -\frac{1}{c^2} \text{curl } \frac{\partial^2 \mathbf{Z}}{\partial t^2}. \quad . . . . . (32)$$

$$\text{Integration yields} \quad \mathbf{E} = -\frac{1}{c^2} \frac{\partial^2 \mathbf{Z}}{\partial t^2} + \text{grad } \phi. \quad . . . . . (33)$$

The arbitrary vector  $\text{grad } \phi$  appears in place of a constant of integration, since the equality of the curls of two vectors implies only that these vectors differ by an irrotational vector  $\text{grad } \phi$ . The two expressions (31) and (33) for  $\mathbf{E}$  will be identical if we take

$$\phi = \text{div } \mathbf{Z} \quad . . . . . (34)$$

and

$$\Delta \mathbf{Z} = \frac{1}{c^2} \frac{\partial^2 \mathbf{Z}}{\partial t^2}. \quad . . . . . (35)$$

Hence, as may easily be verified independently, the values of  $\mathbf{E}$  and  $\mathbf{H}$  defined in terms of  $\mathbf{Z}$  by (30), (33) and (34) satisfy (I), (II) and (III), provided  $\mathbf{Z}$  is a solution of the wave equation (35).

We no longer seek the solution corresponding to a plane wave, for this could give no singularity at the origin; rather, we attempt to find the solution corresponding to a spherical wave, in which the surfaces of equal phase are spheres. For this purpose we write  $\Delta \mathbf{Z}$  in spherical polar co-ordinates, and assume that  $\mathbf{Z}$  is a function of  $r$  only:

$$\frac{\partial^2 \mathbf{Z}}{\partial r^2} + \frac{2}{r} \frac{\partial \mathbf{Z}}{\partial r} = \frac{1}{c^2} \frac{\partial^2 \mathbf{Z}}{\partial t^2}. \quad . . . . . (36)$$

Assuming a simple periodic variation with time, we write

$$\mathbf{Z} = \mathbf{F}(r)e^{i\omega t} \quad . . . . . (37)$$

and obtain the equation

$$\frac{d^2 \mathbf{F}}{dr^2} + \frac{2}{r} \frac{d\mathbf{F}}{dr} + \frac{\omega^2}{c^2} \mathbf{F} = 0. \quad . . . . . (38)$$

for the point function  $\mathbf{F}$ . The integral is

$$\mathbf{F} = \frac{\phi_0}{r} e^{-i\omega r/c}, \quad . . . . . (39)$$

as may be seen by substitution, or by writing the equation in the form

$$\frac{d^2}{dr^2} (r\mathbf{F}) + \frac{\omega^2}{c^2} (r\mathbf{F}) = 0.$$

The negative sign corresponds to a wave spreading outward from the centre  $r = 0$ . We then have

$$\mathbf{Z} = \mathbf{p}_0 \frac{e^{i\omega(t-r/c)}}{r} = \frac{\mathbf{p}(t)}{r} e^{-i\omega r/c}, \quad \dots \quad (40)$$

where  $\mathbf{p}(t)$  replaces  $\mathbf{p}_0 e^{i\omega t}$ . This spherical wave differs from a plane wave in that the amplitude decreases as  $1/r$ , corresponding to the spreading of the energy over a sphere of increasing size. The field vectors  $\mathbf{E}$  and  $\mathbf{H}$  are obtained from  $\mathbf{Z}$  by simple differentiations. Starting with  $\mathbf{H}$ , the simpler of the two, we have

$$\begin{aligned} \text{curl } \mathbf{Z} &= \text{curl} \{ \mathbf{p} f(r) \} = [\text{grad } f(r), \mathbf{p}(t)] \\ &= \frac{df}{dr} [\mathbf{r}_0 \mathbf{p}] = \frac{df}{dr} e^{i\omega t} [\mathbf{r}_0 \mathbf{p}_0], \end{aligned} \quad (41)$$

where  $\mathbf{r}_0$  is a unit vector along  $\mathbf{r}$ . From (40),

$$\frac{df}{dr} = \left( -\frac{i\omega}{cr} - \frac{1}{r^2} \right) e^{-i\omega r/c}. \quad \dots \quad (42)$$

$$\text{Hence} \quad \mathbf{H} = \frac{1}{c} \frac{\partial}{\partial t} \text{curl } \mathbf{Z} = \frac{i\omega}{c} \left( \frac{1}{r^2} + \frac{i\omega}{cr} \right) e^{i\omega(t-r/c)} [\mathbf{p}_0 \mathbf{r}_0]. \quad (43)$$

The vector  $\mathbf{H}$  is thus seen to be composed of two parts, similarly directed but decreasing according to different powers of  $1/r$ . We can locate the regions where one or the other term predominates by writing

$$\frac{1}{r^2} + \frac{i\omega}{cr} = \frac{1}{r} \left( \frac{1}{r} + i \frac{2\pi\nu}{c} \right) = \frac{1}{r} \left( \frac{1}{r} + \frac{2\pi i}{\lambda} \right). \quad \dots \quad (44)$$

The scale of comparison is thus given by the wave-length. In the near zone ( $r \ll \lambda$ ), the first term predominates; in the distant zone ( $r \gg \lambda$ )—the actual radiation field—the second term controls the situation. Between the two is a region where both terms are of the same order of magnitude. We investigate the near zone first. Here we have

$$\mathbf{H}_N = \frac{i\omega}{c} e^{i\omega(t-r/c)} \cdot \frac{1}{r^2} [\mathbf{p}_0 \mathbf{r}_0] = \frac{1}{cr^2} \left[ \frac{\partial \mathbf{p}}{\partial t} \mathbf{r}_0 \right] e^{-i\omega r/c}. \quad (45)$$

If we compare this expression with the Biot-Savart formula for the magnetic field due to an element of a current-carrying conductor (p. 303), we see that, apart from the factor

$$e^{-i\omega r/c},$$



the field in the local region is that due to a current element

$$I ds = \frac{\partial \mathcal{P}}{\partial t} = i\omega \mathcal{P}_0 e^{i\omega t}. \quad . \quad . \quad . \quad . \quad . \quad (46)$$

The factor

$$e^{-i\omega r/c}$$

corresponds to the finite velocity of propagation of the magnetic field, which causes a phase displacement, increasing with distance from the centre, with respect to the instantaneous value of the current element ("retarded field"). In order to gain a conception of such an alternating current element, we consider an electrical dipole whose length and direction are given by  $l$ , and whose ends carry periodically oscillating charges

$$q = \pm q_0 e^{i\omega t}. \quad . \quad . \quad . \quad . \quad . \quad (47)$$

The moment of the dipole will also be a periodic function of the time:

$$\mathcal{P} = ql = q_0 e^{i\omega t} l, \quad . \quad . \quad . \quad . \quad . \quad (47')$$

and we have

$$\frac{\partial \mathcal{P}}{\partial t} = \frac{\partial q}{\partial t} l. \quad . \quad . \quad . \quad . \quad . \quad (48)$$

But  $\partial q / \partial t$  is the current flowing as the charge changes. Our solution thus gives the magnetic field of a dipole whose axis is fixed and whose moment  $\mathcal{P}$  is periodic. Such a dipole is called an *electrical oscillator* or *Hertzian vibrator*. Later we shall find that the electrical field in the near zone also corresponds with the electrostatic field of a dipole of periodically oscillating moment.

A dipole of this kind may be represented by a short linear antenna in which high-frequency oscillatory currents are induced, say by inductively coupling the antenna to an oscillating circuit. In contrast with steady currents, where the impossibility of isolating individual conductor elements made us reject the Biot-Savart law as a fundamental basis, current elements are experimentally attainable when we deal with oscillations. However, the magnetic field of such a segment is given by the Biot-Savart rule only in the immediate neighbourhood of the conductor, and then only to a first approximation. According to the theory of action at a distance, we should use this rule at arbitrarily great distances from the conductor; this would yield results which are entirely contrary to experience, for at great distances (i.e. in the "radiation field"), the second term of (44) controls the situation. By (43) this gives

$$\mathbf{H}_B = -\frac{\omega^2}{c^2} \cdot \frac{1}{r} e^{i\omega(t-r/c)} [\mathcal{P}_0 \mathbf{r}_0] = + \frac{1}{c^2 r} \left[ \frac{\partial^2 \mathcal{P}}{\partial t^2} \mathbf{r}_0 \right] e^{-i\omega r/c}. \quad (49)$$

This magnetic field which decreases as  $1/r$  is no longer proportional to

the first derivative of  $\phi$ —which by equation (48) is proportional to the antenna current—but to its second derivative, i.e. to the *rate of change* of the current.

In order to calculate  $\mathbf{E}$  it is best to start with equation (31) (p. 336):

$$\mathbf{E} = \text{grad div } \mathbf{Z} - \Delta \mathbf{Z}. \quad . \quad . \quad . \quad . \quad (31)$$

We also use the fact that  $\mathbf{Z}$  satisfies the wave equation, i.e.

$$\mathbf{E} = \text{grad div } \mathbf{Z} - \frac{1}{c^2} \frac{\partial^2 \mathbf{Z}}{\partial t^2}. \quad . \quad . \quad . \quad . \quad (31')$$

Now by equation (40) (p. 337),

$$\begin{aligned} \text{div } \mathbf{Z} &= e^{i\omega t} \text{div} \left( \phi_0 \frac{e^{-i\omega r/c}}{r} \right) = e^{i\omega t} \text{div} \{ \phi_0 f(r) \} = e^{i\omega t} \phi_0 \text{grad } f(r) \\ &= e^{i\omega t} \frac{1}{r} \frac{df}{dr} \phi_0 \mathbf{r}, \quad . \quad . \quad . \quad . \quad (50) \end{aligned}$$

whence

$$\text{grad div } \mathbf{Z} = e^{i\omega t} \left\{ \phi_0 \mathbf{r}_0 \left( \frac{d^2 f}{dr^2} - \frac{1}{r} \frac{df}{dr} \right) \cdot \mathbf{r}_0 + \frac{1}{r} \frac{df}{dr} \phi_0 \right\}, \quad (51)$$

where  $\phi_0 \mathbf{r}$  in the first term has been replaced by  $\phi_0 \mathbf{r}_0 r$  in order to make clear which powers of  $r$  occur. We also have the expression for  $df/dr$  given by equation (42) (p. 337):

$$\frac{df}{dr} = \left( -\frac{i\omega}{cr} - \frac{1}{r^2} \right) e^{-i\omega r/c}, \quad . \quad . \quad . \quad . \quad (52)$$

$$\text{and, further,} \quad \frac{d^2 f}{dr^2} = \left( -\frac{\omega^2}{c^2 r} + \frac{2i\omega}{cr^2} + \frac{2}{r^3} \right) e^{-i\omega r/c}. \quad . \quad . \quad . \quad (53)$$

$$\text{We also have} \quad \frac{1}{c^2} \frac{\partial^2 \mathbf{Z}}{\partial t^2} = -\frac{\omega^2}{c^2} \cdot \frac{1}{r} \phi_0 e^{i\omega(t-r/c)}. \quad . \quad . \quad . \quad (54)$$

The field in the near zone is governed by the highest power of  $1/r$  only. Since  $\mathbf{r}_0$  is a unit vector, this is the third. But this power does not occur in the term

$$\frac{1}{c^2} \frac{\partial^2 \mathbf{Z}}{\partial t^2},$$

so that the electric field in the near region is derivable from a scalar potential

$$V_N = -\text{div } \mathbf{Z}. \quad . \quad . \quad . \quad . \quad (55)$$

Neglecting the lower powers of  $1/r$ , this quantity becomes

$$V_N = \frac{\phi_0 \mathbf{r}}{r^3} e^{i\omega(t-r/c)} = \frac{\phi \mathbf{r}_0}{r^2} e^{-i\omega r/c}. \quad . \quad . \quad . \quad (56)$$

But according to equation (28) (p. 273), this is precisely the potential

of a dipole of moment  $\mathbf{p}$ , taking into consideration the variation of phase with increasing distance caused by the finite velocity of propagation. This is then the so-called "retarded potential".

To determine  $\mathbf{E}$  in the radiation field we must collect the terms in  $1/r$ . Since  $\mathbf{r}_0^2 = 1$ , this gives

$$\mathbf{E}_R = \frac{\omega^2}{c^2 r} e^{i\omega(t-r/c)} (\mathbf{p}_0 \cdot \mathbf{r}_0^2 - \mathbf{r}_0 \cdot \mathbf{r}_0 \mathbf{p}_0) = \frac{\omega^2}{c^2 r} e^{i\omega(t-r/c)} [\mathbf{r}_0 [\mathbf{p}_0 \mathbf{r}_0]]. \quad (57)$$

Taking the axis of the dipole to be the polar axis,  $\mathbf{H}$  has the direction of the latitude circles [cf. equation (49)], and  $\mathbf{E}$  is along the meridians [equation (57)] in the far region.  $\mathbf{E}$ ,  $\mathbf{H}$  and the unit vector  $\mathbf{r}_0$  in the direction of propagation again form a right-handed system (fig. 2). In contrast with the phase, the amplitudes of  $\mathbf{E}$  and  $\mathbf{H}$  are not constant over the surface of the sphere, but decrease toward the pole ( $\theta = 0$ ) as  $\sin \theta$ , on account of the factor  $[\mathbf{p}_0 \mathbf{r}_0]$ . This means that a linear oscillator emits no radiation in the

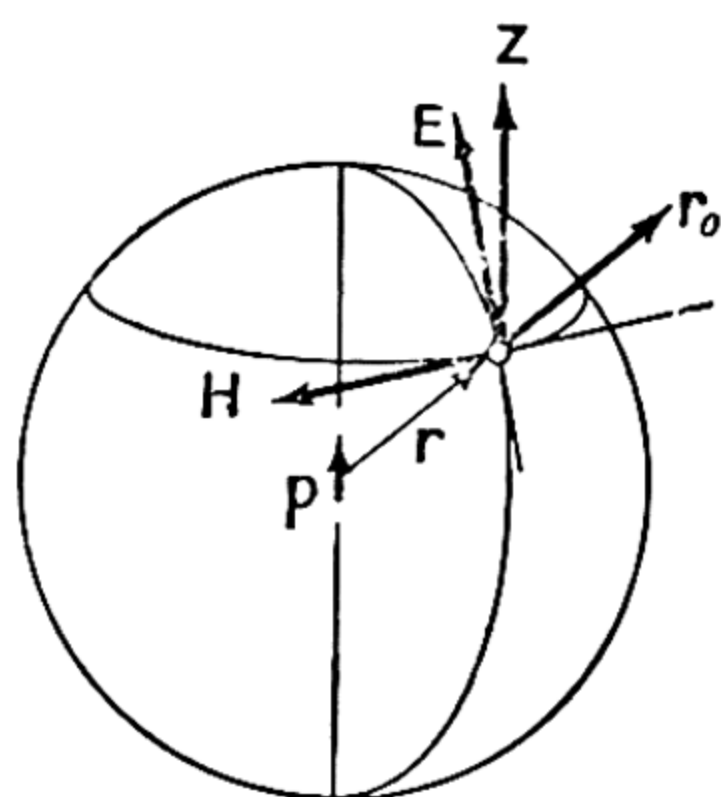


Fig. 2

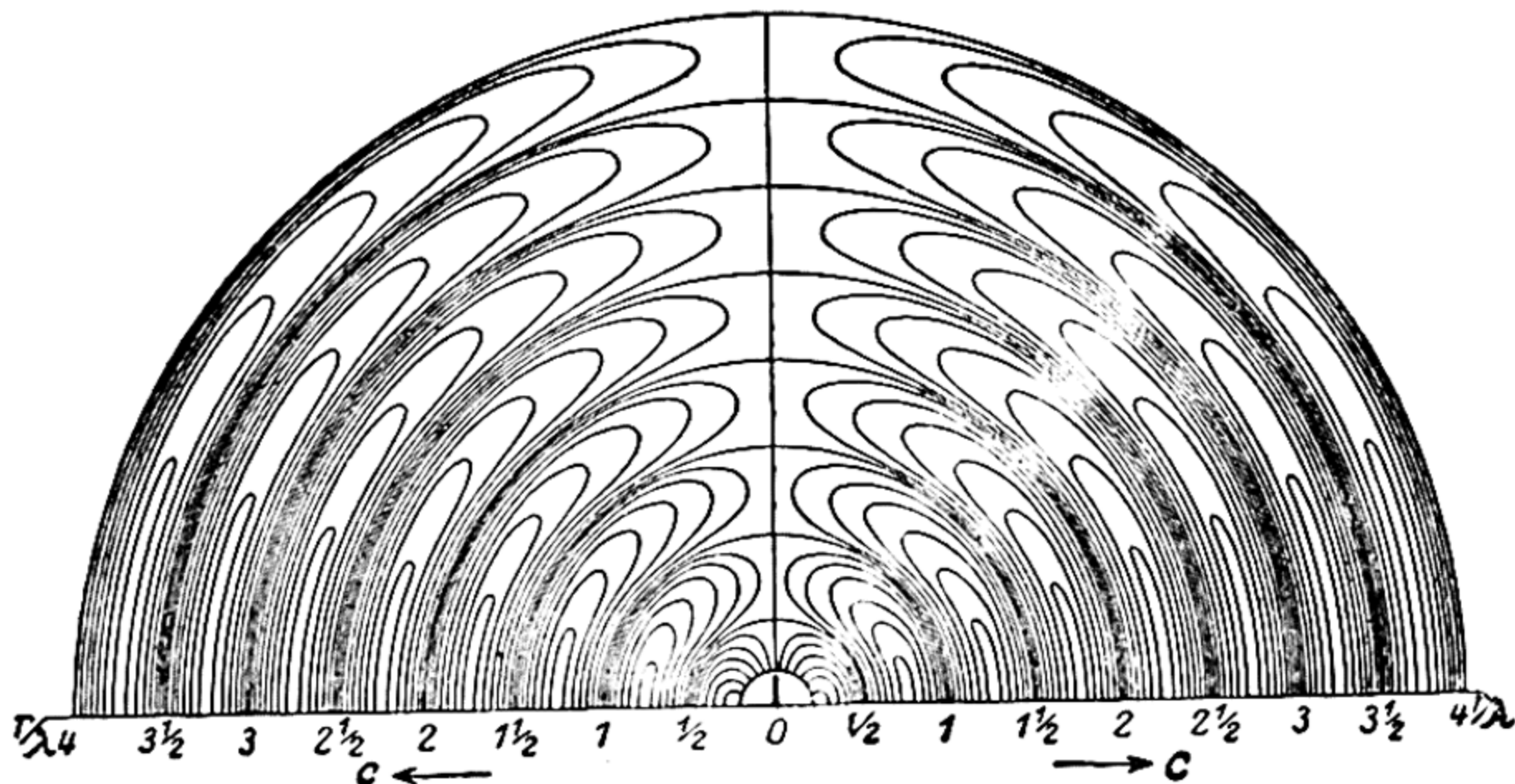


Fig. 3

direction of its axis. This further implies that on account of the vanishing of the radiation term near  $\theta = 0$ , the contribution of the term for the near zone may not be neglected entirely, in spite of its decrease with  $1/r^3$ , since it does not vanish for  $\theta = 0$ . We are therefore not free to draw the meridional direction of  $\mathbf{E}$  quite to the pole. The contribution of the first term causes the lines of force to become rounded at the poles, giving Hertz's celebrated instantaneous picture of the lines of force (fig. 3).\*

\* Consult H. Hertz, *Ann. Phys. Chem.*, 36, p. 1 (1889), or his *Gesammelte Werke*, II, p. 146.



If we now consider the energy flux vector  $\mathbf{S}$ , which is in the direction of  $\mathbf{r}_0$ , its instantaneous value is given by  $c/4\pi$  times the product of the magnitudes of the mutually perpendicular vectors  $\mathbf{E}_R$  and  $\mathbf{H}_R$ . Passing to the real functions, these are

$$|\mathbf{E}_R| = \frac{\omega^2}{c^2 r} p_0 \sin \theta \cos \left\{ \omega \left( t - \frac{r}{c} \right) \right\},$$

$$|\mathbf{H}_R| = \frac{\omega^2}{c^2 r} p_0 \sin \theta \cos \left\{ \omega \left( t - \frac{r}{c} \right) \right\}. \quad \dots \quad (58)$$

Hence we have

$$|\mathbf{S}| = \frac{c}{4\pi} \cdot \frac{\omega^4}{c^4 r^2} p_0^2 \sin^2 \theta \cos^2 \omega \left\{ \left( t - \frac{r}{c} \right) \right\}. \quad \dots \quad (59)$$

Since  $\overline{\cos^2 \phi} = 1/2$ , the time average is,

$$|\bar{\mathbf{S}}| = \frac{c}{8\pi} \frac{\omega^4}{c^4 r^2} p_0^2 \sin^2 \theta. \quad \dots \quad (60)$$

In wireless telegraphy the energy radiated in various directions is conveniently represented by a polar diagram in which the length of the radius vector  $\rho$  is taken equal to the radiated energy, on some suitable scale. The equation of such a curve for the linear oscillator is, from the above

$$\rho = \rho_0 \sin^2 \theta. \quad \dots \quad (61)$$

The average energy which passes through a sphere of radius  $r$  each second is obtained by taking the surface integral

$$\oint \bar{\mathbf{S}} d\mathbf{A} = \frac{c}{8\pi} \frac{\omega^4}{c^4 r^2} p_0^2 \cdot 2\pi r^2 \int_0^\pi \sin^2 \theta \sin \theta d\theta = \frac{c}{8\pi} \cdot \frac{\omega^4}{c^4} 2\pi \cdot \frac{4}{3} p_0^2. \quad (62)$$

This gives the energy lost by the source of radiation each second:

$$\bar{S} = \frac{16\pi^4 c}{3\lambda^4} p_0^2. \quad \dots \quad (63)$$

This expression can be given another form by inserting the current amplitude  $I_0$  in place of the amplitude of the dipole moment. According to equations (46) and (48) (p. 338), the magnitude of  $I_0$  is  $\omega p_0/ds$ . If we put the length  $ds$  of the element (the antenna) equal to  $l$ ,

$$\bar{S} = \frac{4\pi^2}{3c} I_0^2 \left( \frac{l}{\lambda} \right)^2. \quad \dots \quad (64)$$

This loss of energy by radiation is proportional to the square of the current, like the ordinary Joule loss. With an ohmic resistance  $R$ ,

the same alternating current would have an average energy loss in the form of heat amounting to

$$\overline{W} = \frac{RI_0^2}{2}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (65)$$

The ohmic resistance which would cause the same loss of energy as the radiation is called the "radiation resistance"  $R_s$  of the system. According to the above

$$R_s = \frac{8\pi^2}{3c} \left( \frac{l}{\lambda} \right)^2 \text{ electrostatic units,}$$

or

$$80\pi^2 \left( \frac{l}{\lambda} \right)^2 \text{ ohms.} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (66)$$

It is seen that even for antenna lengths of but 1/10 the wave-length, the radiation resistance is of the order of 10 ohms, i.e. large compared with the ohmic resistance.

*Ex. 94.* Show that the phase velocity of  $\mathbf{H}$  is greater than  $c$  in the transitional region between the near and distant zones.

## CHAPTER XVIII

### ELECTROMAGNETIC WAVES: II. PHENOMENA IN TWO ADJOINING MEDIA

#### 1. Unified Rigorous Derivation of the Field Equations and of the Boundary Conditions.

In setting up the electromagnetic field equations we started with common facts of experience. In proceeding thus, we were unable to give a rigorous proof of the electrical analogue of the law of magnetic induction, equation (1) (p. 327); moreover, the boundary conditions were derived only for static or stationary fields. Hence before proceeding further, we shall explain a more satisfactory way of arriving at these results. The proposed method is based upon five hypotheses which, it must be admitted, are not as immediately obtainable from experiment as are those which formed the basis of our previous method. However, the following procedure has the advantage of generality and compactness. The five assumptions are the following:

1. The expression for the density of electrical energy is

$$u_{el} = \frac{K}{8\pi} E^2. \quad . . . . . (1)$$

2. The expression for the magnetic energy density is

$$u_{mag} = \frac{\mu}{8\pi} H^2. \quad . . . . . (2)$$

3. The energy flux is given by

$$S = \frac{c}{4\pi} [EH]. \quad . . . . . (3)$$

4. The expression for the electrical energy converted into Joule heat per second per cubic centimetre is

$$\frac{du_{el}}{dt} = -\sigma E^2. \quad . . . . . (4)$$

5. The experimental fact of the linear superposition of electromagnetic fields, which requires that the field equations be linear and homogeneous in  $E$  and  $H$ .\*

\* We omit from consideration here the special case of ferro-magnetic media, where non-linearity of the equations results from the fact that  $\mu$  itself depends upon the value of the field.



These five postulates emerge in a plausible way from general considerations if we seek the simplest and most direct dimensionally correct relations between the quantities involved. This is essentially what was done on p. 288 in obtaining an expression for the energy density.

The field equations for a uniform medium follow at once from the following energy considerations: the decrease per second of the electromagnetic energy within any volume is equal to the sum of the energy converted into heat and the energy which flows out through the bounding surface, i.e.

$$-\frac{d}{dt} \int \left( \frac{K\mathbf{E}^2}{8\pi} + \frac{\mu\mathbf{H}^2}{8\pi} \right) d\tau = \int \left( -\frac{K}{4\pi} \mathbf{E} \frac{\partial \mathbf{E}}{\partial t} - \frac{\mu}{4\pi} \mathbf{H} \frac{\partial \mathbf{H}}{\partial t} \right) d\tau \\ = \int \sigma \mathbf{E}^2 d\tau + \oint \frac{c}{4\pi} [\mathbf{E}\mathbf{H}] d\mathbf{S}. \quad (5)$$

If the surface integral is converted into a volume integral by means of Gauss's theorem, we have, since  $\text{div}[\mathbf{E}\mathbf{H}] = \mathbf{H} \text{curl} \mathbf{E} - \mathbf{E} \text{curl} \mathbf{H}$ .

$$\int \left( K\mathbf{E} \frac{\partial \mathbf{E}}{\partial t} + \mu\mathbf{H} \frac{\partial \mathbf{H}}{\partial t} + 4\pi\sigma\mathbf{E}^2 + c\mathbf{H} \text{curl} \mathbf{E} - c\mathbf{E} \text{curl} \mathbf{H} \right) d\tau = 0. \quad (6)$$

Since the integral is zero for any arbitrary volume, the integrand must vanish. This is assured, and at the same time Assumption 5 is satisfied, by setting equal to zero the factors by which  $\mathbf{E}$  and  $\mathbf{H}$  are scalarly multiplied. From this it follows that

$$\text{curl} \mathbf{H} = \frac{K}{c} \frac{\partial \mathbf{E}}{\partial t} + \frac{4\pi}{c} \sigma \mathbf{E} = \frac{1}{c} \frac{\partial \mathbf{D}}{\partial t} + \frac{4\pi}{c} \mathbf{i}, \quad (I)$$

$$\text{curl} \mathbf{E} = -\frac{\mu}{c} \frac{\partial \mathbf{H}}{\partial t} = -\frac{1}{c} \frac{\partial \mathbf{B}}{\partial t}. \quad (II)$$

The boundary conditions are obtained from the obvious condition that at the surface of separation of two media the energy leaving the first medium must enter the adjoining one, since for finite conductivity no electrical energy can be dissipated as heat in the mathematical surface of separation. If we take the normal to the interface as  $z$ -axis,

$$\mathbf{S}_z^{(1)} = \frac{c}{4\pi} [\mathbf{E}^{(1)} \mathbf{H}^{(1)}]_z = \mathbf{S}_z^{(2)} = \frac{c}{4\pi} [\mathbf{E}^{(2)} \mathbf{H}^{(2)}]_z, \quad (7)$$

$$\text{or} \quad E_x^{(1)} H_y^{(1)} - H_x^{(1)} E_y^{(1)} = E_x^{(2)} H_y^{(2)} - H_x^{(2)} E_y^{(2)}. \quad (8)$$

This condition is secured by taking for the equations to be satisfied at the bounding surface \*

\* The surface conditions (IV) are usually deduced directly from the fundamental equations (I) and (II) above. We take a small rectangular area having two

$$\left. \begin{aligned} E_x^{(1)} &= E_x^{(2)} & H_x^{(1)} &= H_x^{(2)} \\ E_y^{(1)} &= E_y^{(2)} & H_y^{(1)} &= H_y^{(2)} \end{aligned} \right\} \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad (9)$$

or  $E_{\tan}^{(1)} = E_{\tan}^{(2)} \quad \text{and} \quad H_{\tan}^{(1)} = H_{\tan}^{(2)}; \quad \cdot \quad \cdot \quad (IV)$

or using the concept of surface curl,

$$\text{Curl } \mathbf{E} = [\mu(\mathbf{E}_2 - \mathbf{E}_1)] = 0, \quad \text{Curl } \mathbf{H} = 0. \quad \cdot \quad (IV')$$

The relationships governing the normal components are already fixed by this condition of continuity of the tangential components, since  $\mathbf{E}$  and  $\mathbf{H}$  must, in addition, satisfy the field equations in each of the two media. The general validity of the continuity of the normal components of  $\mathbf{D} = K\mathbf{E}$  and of  $\mathbf{B} = \mu\mathbf{H}$  is readily demonstrated by a simple computation using (IV) and the field equations. This also implies, of course, that the normal components of  $\mathbf{E}$  and  $\mathbf{H}$  themselves suffer discontinuities at the interface.

## 2.\* Waves in the Boundary Layer. The Ground Wave.

The Hertzian solution, which includes the antenna as well as the radiation field, fails to represent actual conditions in this respect, that the transmitter is not suspended in space, but is located at the boundary of two media having quite different electrical properties. We must therefore find two solutions which satisfy the boundary conditions as well as the field equations—one for the upper medium (the air), the other for the earth. The problem, replete with difficulties, was solved in a series of papers by Sommerfeld. We restrict our attention to the simplest case—the demonstration of the possibility of a kind of plane wave having its greatest amplitude at the bounding surface, and hence referred to as a surface wave or boundary layer wave. The consideration of this wave apart from the total phenomenon, it is true, is somewhat arbitrary.

We take the surface of separation to be the  $xy$ -plane of a rectangular system of co-ordinates, the direction of propagation being along the positive  $x$ -axis. For a plane wave in a uniform medium,  $\mathbf{E}$  as well as  $\mathbf{H}$  would be normal to the direction of propagation. But the boundary conditions cannot be fulfilled if we make such an assumption for two media. We therefore generalize the hypothesis in two ways: first we assume the waves to be non-uniform, i.e. the amplitude is to be a function of  $z$ , the distance from the interface. Secondly, we

of its sides normal to the surface, and the other two (which are made much longer) on opposite sides of the surface and parallel to it. On the assumption that there are no surface currents, electric or magnetic, it follows at once from I and II respectively that the tangential component of the magnetic vector and that of the electric vector are continuous across the surface. There is one important case in which surface electric currents are present, viz. the case of a perfect conductor (that is, speaking optically, a perfect reflector); in this case the sole condition is that the tangential component of the electric vector vanishes.

assume that  $\mathbf{H}$ , which we take in the  $y$ -direction, is representable by a purely transverse wave. It will be seen later that  $\mathbf{E}$ , on the contrary, has an  $x$ -component in addition. Thus for  $\mathbf{H}$  our hypothesis is

$$\mathbf{H} = H_y \mathbf{j}, \quad H_y = f(z)e^{i\omega(t-x/v)}. \quad . \quad . \quad . \quad (10)$$

We leave the phase velocity of the wave still undetermined; it is to be expected that this will be determined by the electrical constants of the two media. With the above assumptions (10) concerning  $\mathbf{H}$ , the field equations give for the components of  $\mathbf{E}$ :

$$-\frac{df}{dz} e^{i\omega(t-x/v)} = \frac{K}{c} \frac{\partial E_x}{\partial t} + \frac{4\pi\sigma}{c} E_x. \quad . \quad . \quad . \quad (11)$$

$$0 = \frac{K}{c} \frac{\partial E_y}{\partial t} + \frac{4\pi\sigma}{c} E_y. \quad . \quad . \quad . \quad (12)$$

$$-\frac{i\omega}{v} f e^{i\omega(t-x/v)} = \frac{K}{c} \frac{\partial E_z}{\partial t} + \frac{4\pi\sigma}{c} E_z. \quad . \quad . \quad . \quad (13)$$

We see from (12) that  $E_y$  may be set equal to zero; on the other hand, the  $x$ -component of  $\mathbf{E}$  does not vanish.

The Maxwell equations (II, p. 344) also indicate that  $\mathbf{E}$  is independent of the  $y$ -co-ordinate. The second of them yields the equation

$$\frac{\partial E_x}{\partial z} - \frac{\partial E_z}{\partial x} = -\frac{i\mu\omega}{c} f e^{i\omega(t-x/v)}. \quad . \quad . \quad . \quad (14)$$

We now assume non-uniform waves, propagated in the  $x$ -direction, for  $E_x$  and  $E_z$  also:

$$E_x = g(z)e^{i\omega(t-x/v)}. \quad . \quad . \quad . \quad . \quad (15)$$

$$E_z = h(z)e^{i\omega(t-x/v)}. \quad . \quad . \quad . \quad . \quad (16)$$

Substituting this value of  $E_x$  in equation (11), we have

$$g = \frac{-c \frac{df}{dz}}{4\pi\sigma + K\omega i}. \quad . \quad . \quad . \quad . \quad (17)$$

$$\text{From (13) and (16),} \quad h = \frac{-\frac{i\omega c}{v} f}{4\pi\sigma + K\omega i}. \quad . \quad . \quad . \quad . \quad (18)$$

And finally from (11), (13) and (14),

$$\frac{d^2 f}{dz^2} = \left\{ \omega^2 \left( \frac{1}{v^2} - \frac{K\mu}{c^2} \right) + \frac{4\pi\sigma\mu\omega}{c^2} i \right\} f. \quad . \quad . \quad (19)$$

The solution is

$$f = Ae^{\rho z}, \quad . \quad . \quad . \quad . \quad (20)$$



where 
$$\rho = \pm \sqrt{\omega^2 \left( \frac{1}{v^2} - \frac{K\mu}{c^2} \right) + \frac{4\pi\sigma\mu\omega}{c^2}} i. \quad . \quad . \quad (20')$$

This solution has the same form for both media. But since solutions increasing exponentially with  $z$  have no meaning, we must choose the sign before the radical in such way that for  $z > 0$  (i.e. in air) the real part is negative, while for  $z < 0$  (i.e. in the earth) the sign must be so chosen that the real part is positive. Denoting all quantities pertaining to the upper medium (air) by the subscript 0, the solution for this space becomes

$$H_v^{(0)} = A_0 e^{-\rho_0 z} e^{i\omega(t-x/v_0)}, \quad . \quad . \quad . \quad (21)$$

and in the ground: 
$$H_v = A e^{+\rho z} e^{i\omega(t-x/v)}. \quad . \quad . \quad . \quad (22)$$

Also, by (20'),

$$\left. \begin{aligned} \frac{\omega^2}{v_0^2} - \rho_0^2 &= \frac{K_0\mu_0}{c^2} \omega^2 - \frac{4\pi\sigma_0\mu_0\omega i}{c^2} \\ \frac{\omega^2}{v^2} - \rho^2 &= \frac{K\mu}{c^2} \omega^2 - \frac{4\pi\sigma\mu\omega i}{c^2} \end{aligned} \right\} . \quad . \quad . \quad (23)$$

The boundary conditions expressing continuity of the tangential components of  $\mathbf{E}$  and of  $\mathbf{H}$ , which must hold at every instant, can be fulfilled only if the phase velocity is the same on both sides of the interface, i.e.

$$v = v_0. \quad . \quad . \quad . \quad . \quad . \quad (24)$$

Further, for  $z = 0$ ,  $H_v^{(0)} = H_v$  yields the relationship

$$A = A_0. \quad . \quad . \quad . \quad . \quad . \quad (25)$$

The tangential component of  $\mathbf{E}$  is obtained from equation (15), using (17) and (20) and attending to the remark made above as to the sign of  $\rho$ :

$$E_x^{(0)} = + \frac{c\rho_0 A e^{-\rho_0 z} e^{i\omega(t-x/v)}}{4\pi\sigma_0 + K_0\omega i}; \quad E_x = - \frac{c\rho A e^{+\rho z} e^{i\omega(t-x/v)}}{4\pi\sigma + K\omega i}. \quad (26)$$

The boundary condition thus yields the equation

$$\frac{\rho_0}{4\pi\sigma_0 + K_0\omega i} = - \frac{\rho}{4\pi\sigma + K\omega i}. \quad . \quad . \quad . \quad (27)$$

This equation, taken along with (23) and (24), suffices to determine the remaining three unknowns  $\rho_0$ ,  $\rho$  and  $v$ . If we simplify the right side of equation (23) by introducing the notation

$$\frac{K_0\mu_0\omega^2}{c^2} - \frac{4\pi\sigma_0\mu_0\omega i}{c^2} = a_0 + \beta_0 i = b_0, \quad . \quad . \quad (28)$$

and if we call

$$\frac{4\pi\sigma + K\omega i}{4\pi\sigma_0 + K_0\omega i} = q, \quad \cdot \cdot \cdot \cdot \cdot \cdot (29)$$

we have

$$\left. \begin{aligned} \frac{\omega^2}{v^2} - \rho_0^2 &= b_0 \\ \frac{\omega^2}{v^2} - \rho_0^2 q^2 &= b \end{aligned} \right\} \cdot \cdot \cdot \cdot \cdot \cdot (30)$$

From this,

$$\rho_0 = \pm \sqrt{\frac{b - b_0}{1 - q^2}}, \quad \cdot \cdot \cdot \cdot \cdot \cdot (31)$$

and then from equation (27)

$$\rho = -\rho_0 q. \quad \cdot \cdot \cdot \cdot \cdot \cdot (32)$$

Finally, from (30) and (31),

$$\frac{1}{v} = \pm \frac{1}{\omega} \sqrt{\frac{b - q^2 b_0}{1 - q^2}}. \quad \cdot \cdot \cdot \cdot \cdot \cdot (33)$$

It is to be remembered that all three quantities  $\rho_0$ ,  $\rho$  and  $v$  are complex. By separating the real and imaginary parts of the exponential function it may be shown that the solution has the form

$$H_v^{(0)} = A e^{-r_0 z} e^{-i\omega x} e^{i\omega(t - x/u + s_0 z/\omega)} \quad \cdot \cdot \cdot \cdot \cdot \cdot (34)$$

From this result we see that

1. The amplitude of the waves decreases with the distance  $z$  from the surface of separation.
2. The waves suffer a spatial damping in the  $x$ -direction also.
3. The planes of equal phase are not those for which  $x = \text{const.}$ , but are the oblique planes  $x/u + s_0 z/\omega = \text{const.}$

By substituting the solution in equations (15) and (16) we find further that the  $x$ - and  $z$ -components of  $\mathbf{E}$  show a difference in phase, so that the end point of the vector  $\mathbf{E}$  moves on an ellipse, not along a straight line. The plane of this ellipse, however, is not normal to the direction of propagation, as in the case of elliptically polarized free waves, but contains that direction. The above analysis does not correspond entirely to the general physical problem, inasmuch as we have dealt with plane waves only, without taking the source into consideration. The natural extension of the theory consists in seeking a solution corresponding to the Hertz solution for both media, taking the boundary conditions into account. Sommerfeld's papers deal with this problem, but the details cannot be given here.\*

\* An abridged treatment is given in Frank-Mises, *Die Differential und Integralgleichungen der Mechanik und Physik*, II. Teil, p. 541, Braunschweig, Vieweg & Sohn (1927). Consult also W. H. Eccles, *Wireless Telegraphy and Telephony*, p. 159, Van Nostrand Co., New York (1915).

### 3. Consequences of the Boundary Conditions for Insulating Media. The Optical Laws of Reflection and Refraction.

The considerations of § 2 with regard to boundary-layer waves are not applicable to the short electromagnetic waves of the optical range, for the reason that each radiating atom is so far from boundary surfaces (compared with the wave-length) that it may be considered a free radiator in a uniform medium. However, if a wave sent out from a radiating atom encounters the surface of separation of two different media, part will be reflected and part will be transmitted. We assume the source to be so distant that the wave may be considered plane. We also deal, at first, with wave surfaces of unlimited extent.

It is assumed that the two media show no absorption, so that they are insulators, and the interface is taken to be the  $xy$ -plane of a rectangular system of co-ordinates. Let  $\mathbf{n}$ , the direction of the incident ray, be in the  $xz$ -plane. For the present we do not wish to make any special assumptions regarding the directions  $\mathbf{n}'$  of the reflected ray or  $\mathbf{n}''$  of the refracted ray. Assume that the planes determined by each of these directions and the  $z$ -axis make angles  $\phi'$  and  $\phi''$  respectively with the  $xz$ -plane. Denoting the angles of incidence, reflection and refraction by  $\theta$ ,  $\theta'$  and  $\theta''$  respectively (see also fig. 1, p. 351), we have

$$\left. \begin{aligned} \mathbf{n} &= i \sin \theta + k \cos \theta \\ \mathbf{n}' &= i \sin \theta' \cos \phi' + j \sin \theta' \sin \phi' - k \cos \theta' \\ \mathbf{n}'' &= i \sin \theta'' \cos \phi'' + j \sin \theta'' \sin \phi'' + k \cos \theta'' \end{aligned} \right\}. \quad (35)$$

The electric vectors of the three waves are then

$$\left. \begin{aligned} \mathbf{E} &= \mathbf{A} e^{i\omega(t - \mathbf{n} \cdot \mathbf{r} n_1/c)} = \mathbf{A} e^{i\omega\{t - (x \sin \theta + z \cos \theta) n_1/c\}} \\ \mathbf{E}' &= \mathbf{A}' e^{i\omega'(t - \mathbf{n}' \cdot \mathbf{r} n_1/c) + i\delta'} \\ &= \mathbf{A}' e^{i\omega'\{t - (x \sin \theta' \cos \phi' + y \sin \theta' \sin \phi' - z \cos \theta') n_1/c\} + i\delta'} \\ \mathbf{E}'' &= \mathbf{A}'' e^{i\omega''(t - \mathbf{n}'' \cdot \mathbf{r} n_2/c) + i\delta''} \\ &= \mathbf{A}'' e^{i\omega''\{t - (x \sin \theta'' \cos \phi'' + y \sin \theta'' \sin \phi'' + z \cos \theta'') n_2/c\} + i\delta''} \end{aligned} \right\} \quad (36)$$

while the magnetic vectors are given by

$$\mathbf{H} = \sqrt{\frac{\bar{K}_1}{\mu_1}} [\mathbf{n} \mathbf{E}], \quad \mathbf{H}' = \sqrt{\frac{\bar{K}_1}{\mu_1}} [\mathbf{n}' \mathbf{E}'], \quad \mathbf{H}'' = \sqrt{\frac{\bar{K}_2}{\mu_2}} [\mathbf{n}'' \mathbf{E}'']. \quad (37)$$

At the interface we must have, for any values of  $x$  and  $y$ , at any instant

$$\left. \begin{aligned} E_x + E_x' &= E_x'' \\ E_y + E_y' &= E_y'' \end{aligned} \right\} \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad (38)$$

and

$$\left. \begin{aligned} H_x + H_x' &= H_x'' \\ H_y + H_y' &= H_y'' \end{aligned} \right\} \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad (39)$$



This is possible only if the exponents of  $e$  agree in all three terms. This implies

$$\omega' = \omega'' = \omega, \quad . \quad . \quad . \quad . \quad . \quad (40)$$

i.e. there is no change of frequency for reflection and refraction at a stationary surface of separation of two media. A second consequence is

$$\delta' = \delta'' = 0 \quad \text{or} \quad \pi, \quad . \quad . \quad . \quad . \quad . \quad (41)$$

which means that if any phase change occurs at the interface, it must be  $\pi$ . Further, comparison of the exponents yields

$$\phi' = \phi'' = 0,$$

i.e. both the reflected and the refracted rays lie in the plane of incidence—the plane determined by the incident ray and the normal to the surface; this is the  $xz$ -plane in our case. Finally, we obtain

$$\sin \theta = \sin \theta', \quad . \quad . \quad . \quad . \quad . \quad (42)$$

the law of reflection: *The angle of incidence is equal to the angle of reflection*; also

$$n_1 \sin \theta = n_2 \sin \theta'', \quad . \quad . \quad . \quad . \quad . \quad (43)$$

or 
$$\frac{\sin \theta}{\sin \theta''} = \frac{n_2}{n_1} = n, \quad . \quad . \quad . \quad . \quad . \quad (43')$$

the law of refraction (Snell's Law): *The sine of the angle of incidence is equal to the product of the index of refraction and the sine of the angle of refraction.* (See also p. 371.)

#### 4. Polarization and Intensity Relationships for Reflection and Refraction. Fresnel's Formulæ.

The laws derived in the last section are also obtainable from almost any wave theory of light—in fact, from very elementary considerations. A more severe test of the electromagnetic theory is furnished by the question of the intensities of the various waves. While elementary wave considerations are of no help here, the electromagnetic theory gives the required answer directly from the boundary conditions. The equality of the exponents, which follows from the continuity of the tangential components and which is expressed by the laws of reflection and refraction, yields the relationships (when we assume that  $\mu_2 = \mu_1$ )

$$A_x + A_x' = A_x'', \quad . \quad . \quad . \quad . \quad . \quad (44)$$

$$A_y + A_y' = A_y'', \quad . \quad . \quad . \quad . \quad . \quad (45)$$

$$[n\mathbf{A}]_x + [n'\mathbf{A}']_x = n[n''\mathbf{A}'']_x, \quad . \quad . \quad . \quad . \quad (46)$$

$$[n\mathbf{A}]_y + [n'\mathbf{A}']_y = n[n''\mathbf{A}'']_y, \quad . \quad . \quad . \quad . \quad (47)$$

In component form,

$$[nA]_x = -A_y \cos \theta; [n'A']_x = A_y' \cos \theta; [n''A'']_x = -A_y'' \cos \theta'. \quad (48)$$

$$[nA]_y = -A_x \sin \theta + A_z \cos \theta; [n'A']_y = -A_z' \sin \theta - A_x' \cos \theta; \\ [n''A'']_y = -A_z'' \sin \theta' + A_x'' \cos \theta'. \quad (49)$$

This resolution, however, is not the one appropriate to the physical

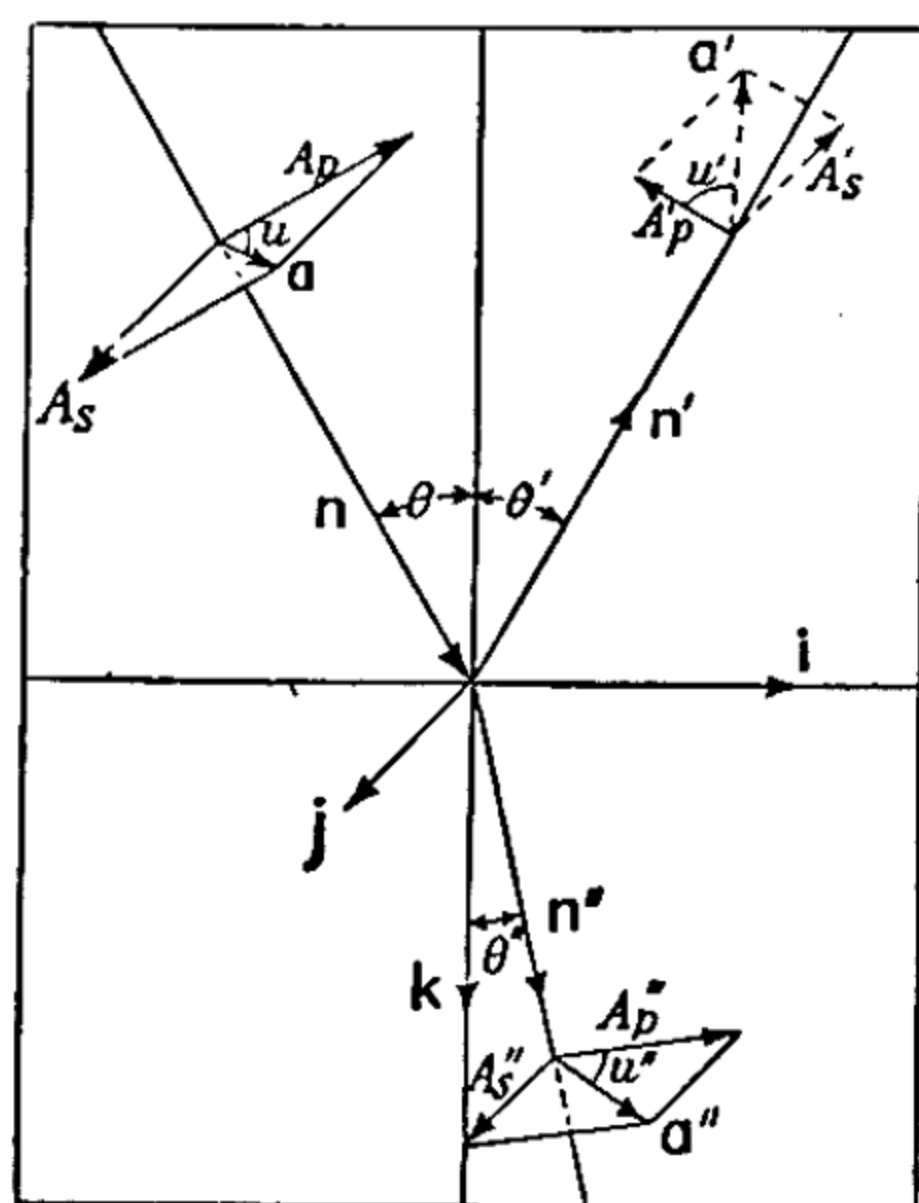


Fig. 1

problem. The natural resolution is that into one component in the plane of incidence and another normal to this plane. In what follows we assume the light to be linearly polarized, i.e. the electric vector remains in a given plane, called the *plane of vibration* of the light. On the other hand, we shall avoid the use of the phrase "plane of polarization", for this term is usually employed for the plane of the magnetic vector, and this is less important for the effects produced (e.g. the photo-electric effect) than is the plane of E. The component of the vibration normal to the plane of incidence is denoted by the subscript s and is given the direc-

tion of the positive y-axis for all three rays:

$$A_s = A_y, \quad A_s' = A_y', \quad A_s'' = A_y''. \quad (50)$$

The positive direction of the components in the plane of incidence, which are characterized by the subscript p, is so chosen that it is obtained by rotating the direction of the ray through  $+90^\circ$  in the plane of incidence. We then have (cf. fig. 1)

$$\left. \begin{aligned} A_x &= A_p \cos \theta \\ A_x' &= -A_p' \cos \theta \\ A_x'' &= A_p'' \cos \theta'' \end{aligned} \right\} \quad (51)$$

$$\left. \begin{aligned} A_z &= -A_p \sin \theta \\ A_z' &= -A_p' \sin \theta \\ A_z'' &= -A_p'' \sin \theta'' \end{aligned} \right\} \quad (52)$$

Substitution in equations (44) to (49) gives

$$(A_p - A_p') \cos \theta = A_p'' \cos \theta''. \quad (53)$$

$$A_s + A_s' = A_s''. \quad (54)$$

$$(A_s - A_s') \cos \theta = n A_s'' \cos \theta'' \quad . \quad . \quad . \quad (55)$$

$$A_p + A_p' = n A_p'' \quad . \quad . \quad . \quad (56)$$

These are four equations for the four unknowns  $A_p'$ ,  $A_s'$ ,  $A_p''$  and  $A_s''$ . Their solution is very simple. We have at once from (53) and (56)

$$A_p' = A_p \frac{n \cos \theta - \cos \theta''}{n \cos \theta + \cos \theta''} = A_p \frac{\sin 2\theta - \sin 2\theta''}{\sin 2\theta + \sin 2\theta''} \quad (57)$$

or 
$$A_p' = A_p \frac{\tan(\theta - \theta'')}{\tan(\theta + \theta'')} \quad . \quad . \quad . \quad (57')$$

Likewise, from equations (54) and (55),

$$A_s' = -A_s \frac{\sin(\theta - \theta'')}{\sin(\theta + \theta'')} \quad . \quad . \quad . \quad (58)$$

Further, by substituting these values in (53) and (55) we have

$$A_p'' = A_p \frac{2 \sin \theta'' \cos \theta}{\sin(\theta + \theta'') \cos(\theta - \theta'')} \quad . \quad . \quad (59)$$

and 
$$A_s'' = A_s \frac{2 \sin \theta'' \cos \theta}{\sin(\theta + \theta'')} \quad . \quad . \quad . \quad (60)$$

The equations (57) to (60) giving the relationships between the amplitudes of the rays are known as Fresnel's formulæ, after their discoverer.

The negative sign in equation (58) has the following meaning: If  $\theta > \theta''$ , i.e. if the ray encounters an "optically denser" medium ( $n > 1$ ),  $A_s$  changes its direction upon being reflected. On the other hand, for  $(\theta + \theta'') < \pi/2$ ,  $A_p'$  has the same sign as  $A_p$ , but the positive direction is now reversed (p. 351), so that both components change sign on reflection from a denser medium, which means that the electric vector has a phase change of  $\pi$ . This jump in phase is of importance in several interference effects that depend on difference of optical path, and must be taken into account by the addition (or subtraction) of half a wave-length.

The ratio of the amplitude of the normal component to that of the component parallel to the incident plane gives the tangent of the azimuth  $u$  of the plane of vibration. Division of equation (58) by equation (57') gives

$$|\tan u'| = \tan u \frac{\cos(\theta - \theta'')}{\cos(\theta + \theta'')} \quad . \quad . \quad . \quad (61)$$

and similarly, from (60) and (59),

$$|\tan u''| = \tan u \cos(\theta - \theta'') \quad . \quad . \quad . \quad (62)$$

From these results certain others follow.



In the first place, (61) shows that for  $\theta + \theta'' = \pi/2$  we always have

$$|\tan u'| = \infty, \quad u' = \frac{\pi}{2}. \quad . \quad . \quad . \quad . \quad (61')$$

Thus, if the reflected and refracted rays are perpendicular to each other, the plane of vibration of the former will be normal to the incident plane; as we may say (see 57'), the component in the plane of incidence will not be reflected at all. If, then, we have ordinary light—i.e. light whose planes of vibration have any random orientations—reflected under these conditions, we obtain plane polarized light whose plane of vibration is normal to the plane of incidence. In the older literature the incident plane in these circumstances is called the plane of polarization, and is the plane of the magnetic vector.

If we combine Snell's Law (p. 350) with the condition  $\theta_p + \theta_p'' = \pi/2$ , we obtain the following relationship for the polarizing angle  $\theta_p$ , i.e. the angle of incidence for which natural light is converted into plane polarized light:

$$\tan \theta_p = n. \quad . \quad . \quad . \quad . \quad . \quad (63)$$

This is *Brewster's Law*. For the reflection of ordinary light at any other angle we obtain partially polarized light, for then the component in the incident plane is also reflected, although to a different degree from the normal component. Since  $\theta$  and  $\theta''$  cannot be greater than  $\pi/2$ , we have for the reflected light

$$|\tan u'| > \tan u, \quad . \quad . \quad . \quad . \quad . \quad (64)$$

i.e. the plane of vibration of plane polarized light is turned *farther* from the plane of incidence upon reflection.

According to (62), we see that the plane of vibration of the refracted ray is turned *toward* the incident plane, but never comes quite into coincidence with it. Nevertheless, since the direction of this turning does not change when the ray emerges from a denser medium— $\cos(\theta - \theta'')$  having the same sign, whichever angle is the greater—we can polarize ordinary light to a high degree by sending it through a number of plane parallel plates separated by air films. This is the so-called "pile of plates" for polarizing light by refraction.

For normal incidence Fresnel's formulæ give indeterminate values, but the solution of the system of equations (53) to (56) (p. 351) is particularly simple in this case. The results are

$$\left. \begin{aligned} A_p' &= A_p \frac{n-1}{n+1} & A_p'' &= \frac{2A_p}{n+1} \\ A_s' &= -A_s \frac{n-1}{n+1} & A_s'' &= \frac{2A_s}{n+1} \end{aligned} \right\} . \quad . \quad . \quad (65)$$

For perpendicular incidence there can be no difference between  $|A_p|$  and  $|A_s|$ , since the plane of incidence is entirely undetermined in this instance. The *reflecting power*  $R$  is defined as the ratio of reflected to incident intensity for normal incidence:

$$R = \frac{I'}{I} = \frac{A'^2}{A^2} = \frac{(n-1)^2}{(n+1)^2}. \quad \cdot \cdot \cdot \cdot (66)$$

*Ex. 95.* The degree of polarization is defined as the ratio

$$P = \frac{I_s - I_p}{I_s + I_p}.$$

What is the value of  $P$  for reflection from glass ( $n = 1.5$ ) at an angle of incidence of  $45^\circ$ ?

*Ex. 96.* Show that a balance of energy flow is maintained for reflection and refraction by an insulator.

## 5. Total Reflection.

The law of refraction (Snell's Law),

$$\sin \theta'' = \frac{1}{n} \sin \theta,$$

leads, when  $\theta$  exceeds a certain value, to values of  $\sin \theta''$  which are greater than unity—i.e. to imaginary values of  $\theta''$ —for light passing from an optically denser medium to one which is less dense ( $n < 1$ ). The elementary theory contents itself with this observation and concludes from it that no light whatsoever passes over into the less dense medium—in other words, the reflection is total. But since we have already used complex functions in our analysis, we need not be disturbed by complex values of  $\theta''$ ; as a matter of fact, we obtain very important results by extending the Fresnel formulæ to include this case. In these formulæ (p. 352) we thus put, if  $\sin \theta > n$ ,

$$\sin \theta'' = \frac{1}{n} \sin \theta, \quad \cos \theta'' = \frac{i}{n} \sqrt{\sin^2 \theta - n^2} \quad \cdot \cdot \cdot (67)$$

and obtain

$$A_p' = A_p \frac{\sin \theta \cos \theta - \sin \theta'' \cos \theta''}{\sin \theta \cos \theta + \sin \theta'' \cos \theta''} = A_p \frac{\cos \theta - \frac{i}{n^2} \sqrt{\sin^2 \theta - n^2}}{\cos \theta + \frac{i}{n^2} \sqrt{\sin^2 \theta - n^2}}. \quad (68)$$

The complex quantity by which  $A_p$  is multiplied is of the form

$$\frac{\rho e^{-i\phi}}{\rho e^{+i\phi}} = e^{-2i\phi}, \quad \cdot \cdot \cdot \cdot (69)$$

where

$$\tan \phi = \frac{\sqrt{\sin^2 \theta - n^2}}{n^2 \cos \theta}. \quad \cdot \cdot \cdot \cdot (70)$$

As appears from (36) (p. 349),

$$E_p' = A_p e^{i\omega(t - n' r n_1/c) - 2i\phi}, \quad . . . . . (71)$$

the component of  $\mathbf{E}$  in the incident plane undergoes a change of phase of  $-2\phi$  for total reflection. Similarly we find

$$A_s' = A_s \frac{\cos \theta - i \sqrt{\sin^2 \theta - n^2}}{\cos \theta + i \sqrt{\sin^2 \theta - n^2}} = A_s e^{-2i\psi}, \quad . . . (72)$$

where

$$\tan \psi = \frac{\sqrt{\sin^2 \theta - n^2}}{\cos \theta}. \quad . . . . . (73)$$

Thus the phase changes are not the same for both components; the components are displaced *relative* to one another. But this means that the originally linearly polarized light, in which the electric vector remains in a given line in a plane normal to the ray, becomes elliptically polarized, i.e. the end point of  $\mathbf{E}$  describes an ellipse. The phase difference  $\delta$  of the components is given, according to the above, by

$$\delta = 2(\phi - \psi). \quad . . . . . (74)$$

In order to determine  $\delta$  we consider the equation

$$\tan \frac{\delta}{2} = \tan(\phi - \psi) = \frac{\tan \phi - \tan \psi}{1 + \tan \phi \tan \psi} = \frac{\sqrt{\sin^2 \theta - n^2} \left( \frac{1}{n^2} - 1 \right)}{\cos \theta \left\{ 1 + \frac{\sin^2 \theta - n^2}{n^2 \cos^2 \theta} \right\}}. \quad (75)$$

After a simple reduction this becomes

$$\tan \frac{\delta}{2} = \frac{\cos \theta \sqrt{\sin^2 \theta - n^2}}{\sin^2 \theta}. \quad . . . . . (76)$$

Since the complex numbers by which  $A_p$  and  $A_s$  are multiplied have the modulus 1, the amplitude is not diminished, and the term "total reflection" is justified.

The results obtained by applying the Fresnel formulæ to the second medium where—according to the elementary theory—nothing should occur, seem quite remarkable at first sight. Since the angle  $\theta''$  becomes complex for values beyond  $\pi/2$ , we set this angle equal to  $\pi/2 + i\beta$  and use the formulæ

$$\cos i\beta = \cosh \beta; \quad \sin i\beta = i \sinh \beta; \quad \cosh^2 \beta - \sinh^2 \beta = 1, \quad (77)$$

i.e.,

$$\sin \theta'' = \cosh \beta = \frac{\sin \theta}{n},$$

$$\cos \theta'' = -i \sinh \beta = -\frac{i}{n} \sqrt{\sin^2 \theta - n^2}. \quad . . . (78)$$



We then obtain

$$\begin{aligned} \mathbf{E}'' &= \mathbf{A}'' e^{i\omega \left( t - \frac{x \cosh \beta}{c/n_2} + \frac{iz \sinh \beta}{c/n_2} \right)} \\ &= \mathbf{A}'' e^{-\frac{\omega z \sinh \beta}{c/n_2}} e^{i\omega \left( t - \frac{x \cosh \beta}{c/n_2} \right)}. \quad \dots (79) \end{aligned}$$

This represents a non-homogeneous wave whose amplitude decreases with extreme rapidity as it penetrates into the second medium. A numerical example for glass and air ( $n = 2/3$ ) shows that for  $\theta = 60^\circ$  the amplitude decreases to  $1/180$  in travelling a distance equal to the wave-length. These waves are propagated along the  $x$ -axis with the phase velocity  $c/(n_2 \cosh \beta)$ . It becomes easier to understand that there is a wave running along the interface in the second medium if we remember that a similar condition arises in the first medium when we have an unlimited beam. The following transformation (cf. p. 67) shows that the superposition of incident and reflected waves results in a condition which may be looked upon as a harmonically modulated wave front travelling along the boundary:

$$\begin{aligned} E_p + E_p' &= A_p e^{i\omega \left( t - \frac{x \sin \theta + z \cos \theta}{c/n_1} \right)} + A_p e^{i\omega \left( t - \frac{x \sin \theta - z \cos \theta}{c/n_1} \right) + i\delta_p} \\ &= 2A_p \cos \left( \frac{\omega z \cos \theta}{c/n_1} + \frac{\delta_p}{2} \right) e^{i\omega \left( t - \frac{x \sin \theta}{c/n_1} \right) + \frac{i\delta_p}{2}}. \quad \dots (80) \end{aligned}$$

The state of affairs in both media may be represented by a diagram of the type shown in fig. 2, in which the phenomenon in the second medium is represented as a continuation of the wave in the first medium. With regard to experimental attempts to verify the existence of the "layer of light" in the second medium, the objection might be raised that the wave would be disturbed by any such experiment, and hence the reflection in the first medium would no longer be total. Nevertheless, it would be justifiable to extrapolate from a very small disturbance in the second medium to no disturbance whatsoever. As long as no disturbance is present there will, on the average, be no transfer of energy into the second medium; for if the Poynting vector  $\mathbf{S}$  is calculated it is found that at any instant there are places where energy enters alternating with places where energy leaves, and that these locations themselves change periodically.

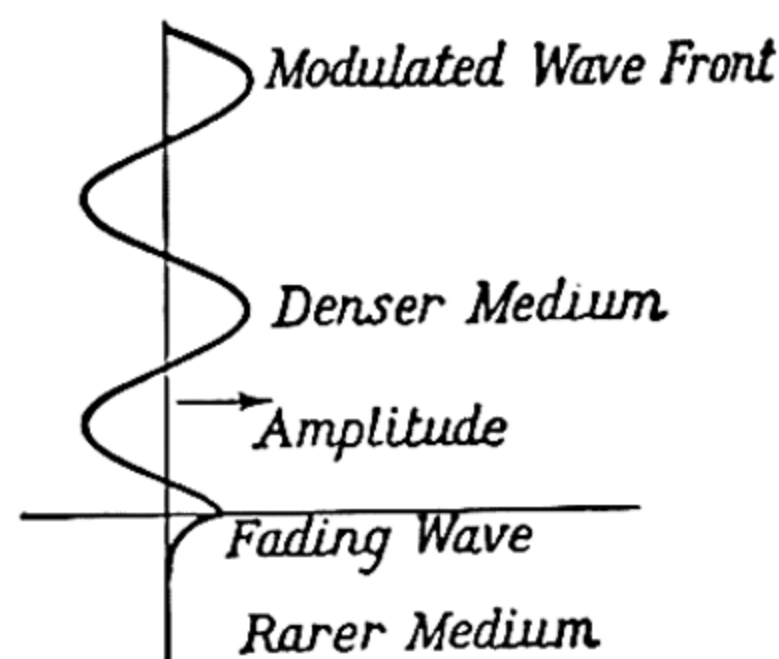


Fig. 2

*Ex. 97.* What index of refraction with respect to air must a substance have in order that circularly polarized light be obtained by one total reflection? At what angle must light be incident on the surface of separation of glass and air ( $n = 2/3$ ) in order to obtain circularly polarized light by two total reflections?

## 6. Absorbing Media. The Optics of Metals.

In considering the propagation of plane waves in conducting media (Chap. XVII, § 4, p. 333) it proved necessary to introduce a complex index of refraction, the quantities  $n$  and  $\kappa$  being functions of  $K$  and  $\sigma$ . For optical frequencies, however, we are not permitted to use the statical values, since these quantities vary considerably with frequency for short waves. For this reason we characterize a conductor by  $n$  and  $\kappa$  instead of by  $K$  and  $\sigma$  in the optical region. If desired, these quantities might be calculated from  $n$  and  $\kappa$ , using formula (23) (p. 334), but this would yield nothing essentially new for the optical region. The connexion with the electrical values does become effective, however, at the long infra-red frequencies.

We must thus substitute

$$\sin \theta'' = \frac{\sin \theta}{n(1 - i\kappa)}. \quad \dots \dots \dots (81)$$

$$\cos \theta'' = \frac{1}{n(1 - i\kappa)} \sqrt{n^2(1 - i\kappa)^2 - \sin^2 \theta}. \quad \dots \dots (82)$$

in the Fresnel formulæ. The ratio  $A_s'/A_p'$  is again complex, i.e. there is again a phase difference between the components in and perpendicular to the plane of incidence, as in the case of total reflection. The reflected light is elliptically polarized.

We thus write

$$\frac{A_s'}{A_p'} = - \frac{A_s}{A_p} \frac{\cos(\theta - \theta'')}{\cos(\theta + \theta'')} = - \frac{A_s}{A_p} \rho e^{i\delta}. \quad \dots \dots (83)$$

If we remove the difference in phase by inserting a compensator (a suitably cut crystal plate) which produces the opposite phase difference, then the azimuth of the restored plane of vibration becomes, according to (83),

$$|\tan u'| = \left| \frac{A_s'}{A_p'} \right| = \rho \frac{A_s}{A_p}. \quad \dots \dots \dots (84)$$

By writing the complex number

$$\frac{\cos(\theta - \theta'')}{\cos(\theta + \theta'')}$$

in the form  $\rho e^{i\delta}$  we can find the phase difference  $\delta$  as well as the real quantity  $\rho$  of (83) for any angle of incidence; this determines the ellipse of vibration. In general, this ellipse will be placed in some arbitrary position in the plane normal to  $n'$ . There is, however, a particular angle—called the *principal angle of incidence*  $\theta^*$  for which the ellipse is so placed that one axis falls in the plane of incidence. In

this case the phase difference between  $A_s'$  and  $A_p'$  is exactly  $\pi/2$ . The azimuth of the plane of vibration restored by inserting a compensator which nullifies the phase difference  $\pi/2$  (a so-called quarter-wave plate, usually made of mica) is called the principal azimuth  $u'^*$ , the azimuth of the incident light being assumed to be that of the plane at  $45^\circ$ . Hence

$$-\frac{A_s'}{A_p'} = i\rho = + \frac{1 + \tan \theta^* \tan \theta^{*''}}{1 - \tan \theta^* \tan \theta^{*''}} \quad \cdot \cdot \cdot (85)$$

It follows from this that

$$\tan \theta^* \tan \theta^{*''} = \frac{\rho + i}{\rho - i} = e^{2i\chi}, \quad \cdot \cdot \cdot (86)$$

where

$$\tan \chi = \frac{1}{\rho} = \cot u'^*$$

(cf. equation (84), p. 357), and hence

$$\chi = \frac{\pi}{2} - u'^*. \quad \cdot \cdot \cdot \cdot \cdot (87)$$

The auxiliary angle  $\chi$  introduced here as an aid to calculation is thus identical with the complement of the principal azimuth  $u'^*$ . If we substitute the values of  $\sin \theta^{*''}$  and  $\cos \theta^{*''}$  from equations (81) and (82) in the expression  $\tan \theta^* \tan \theta^{*''}$ , we have

$$\frac{\tan \theta^* \sin \theta^*}{\sqrt{n^2(1 - i\kappa)^2 - \sin^2 \theta^*}} = e^{2i\chi}. \quad \cdot \cdot \cdot \cdot (88)$$

Separation of real and imaginary parts gives two equations for the unknowns  $n$  and  $\kappa$  in terms of the observable quantities  $\theta^*$  and  $u'^*$ . The actual values of  $n$  and  $\kappa$  are such that  $|(n - i\kappa)^2| \gg \sin^2 \theta^*$ . Hence, if we neglect  $\sin^2 \theta^*$  under the radical, we obtain very simple approximate formulæ which are sufficiently accurate for most purposes. Thus

$$\frac{\tan \theta^* \sin \theta^*}{n(1 - i\kappa)} = e^{2i\chi}. \quad \cdot \cdot \cdot \cdot \cdot (89)$$

Comparing moduli and amplitudes of the complex numbers, we obtain

$$\left. \begin{aligned} \tan \theta^* \sin \theta^* &= n \sqrt{1 + \kappa^2} \\ \kappa &= \tan 2\chi \end{aligned} \right\} \cdot \cdot \cdot \cdot (90)$$

It follows that

$$\left. \begin{aligned} n &= \tan \theta^* \sin \theta^* \cos 2\chi \\ \kappa &= \tan 2\chi \end{aligned} \right\} \cdot \cdot \cdot \cdot (91)$$



From equation (65) (p. 353) the reflecting power of metals is seen to be determined by

$$\frac{A'}{A} = \frac{n - 1 - in\kappa}{n + 1 - in\kappa} = \frac{ae^{i\alpha}}{be^{i\beta}} = \frac{a}{b} e^{i(\alpha - \beta)}. \quad (92)$$

The intensity depends only upon the ratio of the squares of the amplitudes. The reflecting power is thus given by

$$R = \left(\frac{a}{b}\right)^2 = \frac{(n - 1)^2 + n^2\kappa^2}{(n + 1)^2 + n^2\kappa^2} = \frac{n^2 + n^2\kappa^2 + 1 - 2n}{n^2 + n^2\kappa^2 + 1 + 2n}. \quad (93)$$

The greater the value of  $\kappa$ , the nearer the value of this fraction approaches to unity. Hence all wave-lengths which are strongly absorbed are strongly reflected. Hence the colours of such substances by transmitted and by reflected light are complementary, and a thin film of gold, for example, appears blue by transmitted light.

If a plane wave is incident obliquely on the surface separating a metal from a vacuum, the refracted wave is inhomogeneous, i.e. the planes of equal amplitude are no longer identical with the planes of equal phase. This is the case because the various parts of the wave front (which is a plane of constant phase) have traversed different thicknesses of the absorbing material. Reviewing the derivation of the law of refraction (p. 349), one sees that this law must hold even when the index is complex; indeed, the Fresnel formulæ, which involve the refractive index, were applied to metals. On the other hand, if we examine the sine relation without troubling about absorption, a lengthy calculation shows that the ratio of sines depends on the angle between the planes of equal phase and the planes of equal amplitude, and so must depend on the angle of incidence. Hence for the refraction itself the familiar law no longer holds. Because of the strong absorption, however, the refracted waves are only slightly inhomogeneous, so that this remarkable phenomenon is very inconsequential in practice.

## CHAPTER XIX

### ELECTROMAGNETIC WAVES. III: PROPAGATION IN ANISOTROPIC MEDIA. THE OPTICS OF CRYSTALS

#### 1. The Field Equations for Anisotropic Bodies.

For anisotropic media—we have in view chiefly crystals which do not belong to the regular system—the relationships are more complex in that the polarizability of the atoms cannot be expressed by a simple scalar, as in the case of a free atom, for which this quantity averaged over all orientations is the same in all directions. Since it is part of a crystal lattice, the atom is no longer perfectly free to turn and hence it is no longer possible to average over all orientations. Also, in addition to that due to the character of the atom itself, a further element of anisotropy may be introduced by the presence of neighbouring atoms. If the polarization is different for the various components of  $\mathbf{E}$ , the vector  $\mathbf{P}$  no longer has the direction of  $\mathbf{E}$ . One thing, however, must be preserved: the linear connexion between  $\mathbf{P}$  and  $\mathbf{E}$ .  $\mathbf{P}$  thus becomes a linear vector function of  $\mathbf{E}$ , and hence the susceptibility is a tensor:

$$\mathbf{P} = \Psi \mathbf{E}. \quad . . . . . (1)$$

But the vector of the dielectric displacement  $\mathbf{D}$  also becomes a linear function of  $\mathbf{E}$ , and the dielectric constant is now a tensor:

$$\mathbf{D} = \mathbf{E} + 4\pi\mathbf{P} = \mathbf{E} + 4\pi\Psi\mathbf{E} = \Phi\mathbf{E}. \quad . . . (2)$$

In component form,

$$\left. \begin{aligned} D_x &= K_{11}E_x + K_{12}E_y + K_{13}E_z \\ D_y &= K_{21}E_x + K_{22}E_y + K_{23}E_z \\ D_z &= K_{31}E_x + K_{32}E_y + K_{33}E_z \end{aligned} \right\} . . . (3)$$

Since we deal principally with transparent crystals, ferromagnetic bodies, which at the same time possess metallic conductivity, do not come into consideration, and we can set the permeability equal to unity. With this stipulation, the second field equation remains unchanged. How must the first field equation be written for the present case? Since every change of polarization means a motion of electric

charges, or a current, of density  $\partial \mathbf{P} / \partial t$ , we have in general for all media

$$\text{curl } \mathbf{H} = \frac{1}{c} \frac{\partial \mathbf{E}}{\partial t} + \frac{4\pi}{c} \frac{\partial \mathbf{P}}{\partial t} = \frac{1}{c} \frac{\partial \mathbf{D}}{\partial t}. \quad . . . (4)$$

If we consider the derivation of the field equations in Chap. XVIII, § 1 (p. 343), which was based upon perfectly general principles, we see that in order to obtain the form (4) we must make the following assumption for the energy density in the general case, i.e. for anisotropic media:

$$u_{el} = \frac{1}{8\pi} \mathbf{E} \mathbf{D}. \quad . . . . . (5)$$

The energy flux vector, which contains only  $\mathbf{E}$  and  $\mathbf{H}$ , continues to be given by

$$\frac{c}{4\pi} [\mathbf{E} \mathbf{H}].$$

The conservation of energy is thus expressed by

$$\oint \frac{c}{4\pi} [\mathbf{E} \mathbf{H}] d\mathbf{S} = - \frac{\partial}{\partial t} \frac{1}{8\pi} \int (\mathbf{E} \mathbf{D} + \mathbf{H}^2) d\tau, \quad . . (6)$$

or

$$\begin{aligned} \frac{c}{4\pi} \int (\mathbf{H} \text{ curl } \mathbf{E} - \mathbf{E} \text{ curl } \mathbf{H}) d\tau &= - \frac{\partial}{\partial t} \frac{1}{8\pi} \int (\mathbf{E} \mathbf{D} + \mathbf{H}^2) d\tau \\ &= - \frac{1}{8\pi} \int \left( \mathbf{E} \frac{\partial \mathbf{D}}{\partial t} + \mathbf{D} \frac{\partial \mathbf{E}}{\partial t} + 2\mathbf{H} \frac{\partial \mathbf{H}}{\partial t} \right) d\tau. \quad . . (7) \end{aligned}$$

We again obtain the field equations as on p. 344 by equating separately the factors by which  $\mathbf{E}$  and  $\mathbf{H}$  are multiplied. For the second equation this offers no difficulty, but the first equation in the form (4) results only if

$$\mathbf{E} \frac{\partial \mathbf{D}}{\partial t} = \mathbf{D} \frac{\partial \mathbf{E}}{\partial t}. \quad . . . . . (8)$$

But this means that  $\Phi$  is a symmetric tensor, for if the component form of the condition, viz.

$$\left. \begin{aligned} &K_{11}E_x \frac{\partial E_x}{\partial t} + K_{12}E_x \frac{\partial E_y}{\partial t} + K_{13}E_x \frac{\partial E_z}{\partial t} \\ &+ K_{21}E_y \frac{\partial E_x}{\partial t} + K_{22}E_y \frac{\partial E_y}{\partial t} + K_{23}E_y \frac{\partial E_z}{\partial t} \\ &+ K_{31}E_z \frac{\partial E_x}{\partial t} + K_{32}E_z \frac{\partial E_y}{\partial t} + K_{33}E_z \frac{\partial E_z}{\partial t} \end{aligned} \right\}$$



$$= \left\{ \begin{array}{l} K_{11}E_x \frac{\partial E_x}{\partial t} + K_{12}E_y \frac{\partial E_x}{\partial t} + K_{13}E_z \frac{\partial E_x}{\partial t} \\ + K_{21}E_x \frac{\partial E_y}{\partial t} + K_{22}E_y \frac{\partial E_y}{\partial t} + K_{23}E_z \frac{\partial E_y}{\partial t} \\ + K_{31}E_x \frac{\partial E_z}{\partial t} + K_{32}E_y \frac{\partial E_z}{\partial t} + K_{33}E_z \frac{\partial E_z}{\partial t} \end{array} \right\} \quad (8')$$

is to be valid for all values of  $E_x$ ,  $E_y$  and  $E_z$ , we must have

$$K_{ik} = K_{ki}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (9)$$

In this case the connexion between  $\mathbf{D}$  and  $\mathbf{E}$  may be visualized by means of the tensor ellipsoid. If we put  $\mathbf{E} = \mathbf{r}^*$  and draw the surface,

$$\mathbf{D}\mathbf{r} = K_{11}x^2 + 2K_{12}xy + K_{22}y^2 + 2K_{23}yz + K_{33}z^2 + 2K_{31}zx = 1, \quad (10)$$

then, according to p. 37, the direction of  $\mathbf{D}$  is given by the normals to the surface, while its magnitude is the reciprocal of the projection of  $\mathbf{r}$  on the normal. In what follows we shall employ only the co-ordinate system of the principal axes, in which system the equation of the ellipsoid is

$$K_I x'^2 + K_{II} y'^2 + K_{III} z'^2 = 1. \quad . \quad . \quad . \quad (11)$$

$K_I$ ,  $K_{II}$ ,  $K_{III}$  are called the *principal dielectric constants of the crystal*. The axes of the ellipsoid are proportional to the quantities

$$\frac{1}{\sqrt{K_I}}, \quad \frac{1}{\sqrt{K_{II}}}, \quad \frac{1}{\sqrt{K_{III}}}.$$

Now, in an isotropic body, the velocity of light is  $c/\sqrt{K}$ . Similarly, we call the quantities

$$\frac{c}{\sqrt{K_I}}, \quad \frac{c}{\sqrt{K_{II}}}, \quad \frac{c}{\sqrt{K_{III}}},$$

the principal velocities of light in the crystal. The axes of this so-called Fresnel ellipsoid are thus proportional to the principal velocities of light. In the system of the principal axes we have the simple relationships:

$$D_{x'} = K_I E_{x'}, \quad D_{y'} = K_{II} E_{y'}, \quad D_{z'} = K_{III} E_{z'}, \quad . \quad (12)$$

However, since  $K_I$ ,  $K_{II}$  and  $K_{III}$  are unequal, the resulting vector  $\mathbf{D}$  will no longer be parallel to  $\mathbf{E}$ .

\* Of course, this relationship must contain some dimensional factor, since we cannot set an electrical field strength equal to a length. Such a factor, of magnitude unity, is assumed to be contained in this equation.

In this same co-ordinate system we also have, by equation (12):

$$E_x = \frac{1}{K_I} D_x, \quad E_y = \frac{1}{K_{II}} D_y, \quad E_z = \frac{1}{K_{III}} D_z, \quad . \quad (13)$$

i.e. the principal axes of the Fresnel ellipsoid are also the principal axes of that tensor ellipsoid which, reciprocally, gives  $\mathbf{E}$  as a function of  $\mathbf{D}$ . Its equation is

$$\frac{x'^2}{K_I} + \frac{y'^2}{K_{II}} + \frac{z'^2}{K_{III}} = 1. \quad . \quad . \quad . \quad . \quad (14)$$

Thus the axes are proportional to the square roots of the principal dielectric constants—that is, the principal indices of refraction. The figure is thus called the *index ellipsoid*. The corresponding tensor is denoted by  $\Phi^{-1}$ , and so

$$\mathbf{E} = \Phi^{-1} \mathbf{D}. \quad . \quad . \quad . \quad . \quad . \quad (15)$$

## 2. Plane Electromagnetic Waves in Anisotropic Media.

We assume the electric vector to be representable by a plane wave whose planes of equal phase are taken normal to the unit vector  $\mathbf{n}$ . Let the phase be propagated in the direction of  $\mathbf{n}$  with the velocity  $v$ . Thus we have

$$\mathbf{E} = \mathbf{E}_0 e^{i\omega(t - \frac{r\mathbf{n}}{v})}. \quad . \quad . \quad . \quad . \quad . \quad (16)$$

We then have, from the second field equation,

$$\frac{\partial \mathbf{H}}{\partial t} = -c \operatorname{curl} \mathbf{E}_0 e^{i\omega(t - \frac{r\mathbf{n}}{v})} = c [\mathbf{E}_0, \operatorname{grad} e^{i\omega(t - \frac{r\mathbf{n}}{v})}]. \quad . \quad (17)$$

Since  $\operatorname{grad} r\mathbf{n} = \mathbf{n}$ , we obtain after integration with respect to  $t$ ,\*

$$\mathbf{H} = \frac{c}{v} e^{i\omega(t - \frac{r\mathbf{n}}{v})} [\mathbf{n} \mathbf{E}_0] = \mathbf{H}_0 e^{i\omega(t - \frac{r\mathbf{n}}{v})}, \quad . \quad . \quad (18)$$

and

$$\mathbf{H}_0 = \frac{c}{v} [\mathbf{n} \mathbf{E}_0]. \quad . \quad . \quad . \quad . \quad . \quad (19)$$

Thus the magnetic vector is normal to the plane of  $\mathbf{E}$  and  $\mathbf{n}$ .

\* For if we put  $r\mathbf{n} = u$ ,  $\operatorname{grad} u$  becomes equal to  $\mathbf{n}$  and

$$\operatorname{grad} e^{i\omega(t - u/v)} = -\frac{i\omega}{v} e^{i\omega(t - u/v)} \operatorname{grad} u,$$

i.e. 
$$[\mathbf{E}_0, \operatorname{grad} e^{i\omega(t - r\mathbf{n}/v)}] = \frac{i\omega}{v} e^{i\omega(t - r\mathbf{n}/v)} [\mathbf{n} \mathbf{E}_0].$$

However, we do not yet have any information regarding the angle between  $\mathbf{E}$  and  $\mathbf{n}$ . From the first field equation,

$$\frac{\partial \mathbf{D}}{\partial t} = c \operatorname{curl} \mathbf{H} = c \operatorname{curl} \mathbf{H}_0 e^{i\omega(t - \frac{r}{v})} = -\frac{i\omega c}{v} e^{i\omega(t - \frac{r}{v})} [\mathbf{n} \mathbf{H}_0]. \quad (20)$$

Integrating, and inserting (19), we obtain

$$\mathbf{D} = -\frac{c^2}{v^2} e^{i\omega(t - \frac{r}{v})} [\mathbf{n} [\mathbf{n} \mathbf{E}_0]] = \mathbf{D}_0 e^{i\omega(t - \frac{r}{v})}. \quad (21)$$

This signifies that  $\mathbf{D}$  is normal to  $\mathbf{n}$ . With respect to  $\mathbf{D}$  and  $\mathbf{H}$  we thus have pure transverse waves. From (21) we obtain an equation which is of fundamental importance in what follows: \*

$$\left. \begin{aligned} \mathbf{D} &= \frac{c^2}{v^2} (\mathbf{E} - \mathbf{n} \cdot \mathbf{n} \mathbf{E}), \\ \text{or } \frac{v^2}{c^2} \mathbf{D} - \mathbf{E} + \mathbf{n} \cdot \mathbf{n} \mathbf{E} &= 0 \end{aligned} \right\} \dots \dots \dots (22)$$

This equation states that  $\mathbf{D}$ ,  $\mathbf{E}$  and  $\mathbf{n}$  lie in the same plane— $\mathbf{H}$  being normal to this plane, as we saw above. But the vector

$$\mathbf{S} = \frac{c}{4\pi} [\mathbf{E} \mathbf{H}]$$

also lies in this plane; moreover,  $\mathbf{S}$  forms the same angle  $\theta$  with  $\mathbf{n}$  as does  $\mathbf{E}$  with  $\mathbf{D}$  (cf. fig. 1).  $\mathbf{S}$  gives the direction of energy flow; this means that a bounded portion of the wave front travels in this direction while its normal continues to have the direction of  $\mathbf{n}$ . The phenomenon is to be compared with a company of soldiers executing a "right oblique". As may be seen directly from the figure, the velocity of the energy flow is

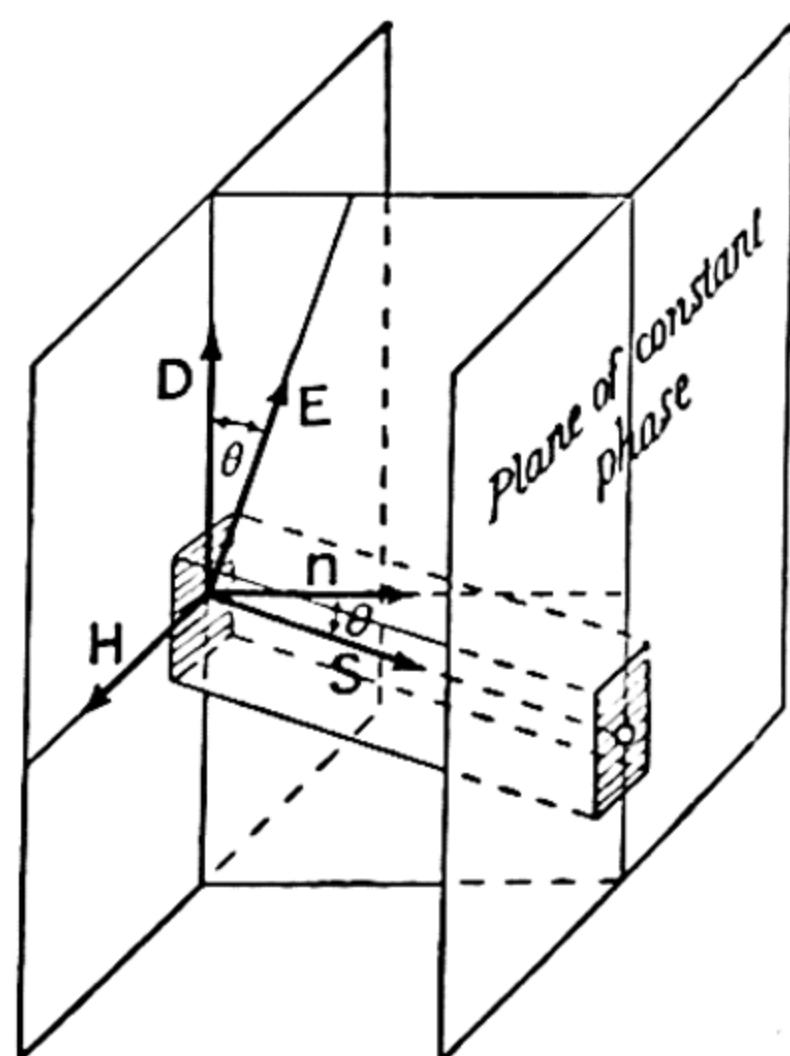


Fig. 1

$$V = \frac{v}{\cos \theta}. \quad \dots \dots \dots (23)$$

$V$  is also called the *ray velocity*.

If we decompose (22) into components referred to the principal system of axes, with which we suppose  $\mathbf{n}$  to make angles

\* The subscript 0 is now omitted, since the equation is true for the instantaneous values  $\mathbf{E}$ ,  $\mathbf{H}$  and  $\mathbf{D}$  as well as for the amplitudes  $\mathbf{E}_0$ ,  $\mathbf{H}_0$  and  $\mathbf{D}_0$ .



$\alpha, \beta, \gamma$ ; and if we express  $\mathbf{E}$  in terms of  $\mathbf{D}$  in the second term, then

$$\left. \begin{aligned} \left( \frac{1}{K_I} - \frac{v^2}{c^2} \right) D_x &= nE \cos \alpha \quad \text{or} \quad D_x = \frac{nE \cos \alpha}{\frac{1}{K_I} - \frac{v^2}{c^2}} \\ \text{Similarly,} \quad D_y &= \frac{nE \cos \beta}{\frac{1}{K_{II}} - \frac{v^2}{c^2}} \\ D_z &= \frac{nE \cos \gamma}{\frac{1}{K_{III}} - \frac{v^2}{c^2}} \end{aligned} \right\} \quad (24)$$

The scalar product  $\mathbf{D}n$  vanishes, since  $\mathbf{D}$  is perpendicular to  $n$ . If we express this product in terms of the components given by (24), and if we divide by the non-vanishing product  $\mathbf{E}n$ , we obtain the relationship

$$\frac{\cos^2 \alpha}{\frac{1}{K_I} - \frac{v^2}{c^2}} + \frac{\cos^2 \beta}{\frac{1}{K_{II}} - \frac{v^2}{c^2}} + \frac{\cos^2 \gamma}{\frac{1}{K_{III}} - \frac{v^2}{c^2}} = 0, \quad (25)$$

This is a quadratic equation for  $v^2$  in terms of the components of  $n$ ; hence to every wave normal there correspond, in general, two phase velocities. For the direction of the  $x$ -axis, for example, we have  $\cos \alpha = 1, \cos \beta = \cos \gamma = 0$ . Hence

$$v'^2 = \frac{c^2}{K_{II}} \quad \text{and} \quad v''^2 = \frac{c^2}{K_{III}}.$$

We may therefore replace  $c/\sqrt{K_I}, \dots$  by the principal velocities  $v_I, v_{II}$  and  $v_{III}$ , and thus obtain the more elegant form

$$\frac{\cos^2 \alpha}{v_I^2 - v^2} + \frac{\cos^2 \beta}{v_{II}^2 - v^2} + \frac{\cos^2 \gamma}{v_{III}^2 - v^2} = 0, \quad (25')$$

If we know the values of  $v$  corresponding to a given direction—call them  $v'$  and  $v''$ —we obtain the directions of the corresponding vectors  $D'$  or  $D''$  by substituting  $v'$  or  $v''$  in the following equations which come from (24):

$$D_x : D_y : D_z = \frac{\cos \alpha}{v_I^2 - v^2} : \frac{\cos \beta}{v_{II}^2 - v^2} : \frac{\cos \gamma}{v_{III}^2 - v^2}. \quad (26)$$

These two directions are mutually perpendicular, for we have

$$\mathbf{D}'\mathbf{D}'' = (\mathbf{nE}')(\mathbf{nE}'')c^4 \left\{ \frac{\cos^2 \alpha}{(v_I^2 - v'^2)(v_I^2 - v''^2)} + \frac{\cos^2 \beta}{(v_{II}^2 - v'^2)(v_{II}^2 - v''^2)} + \frac{\cos^2 \gamma}{(v_{III}^2 - v'^2)(v_{III}^2 - v''^2)} \right\} \quad (27)$$

Now we also have the identity

$$\frac{\cos^2 \alpha}{v_I^2 - v'^2} - \frac{\cos^2 \alpha}{v_I^2 - v''^2} \equiv \frac{(v'^2 - v''^2) \cos^2 \alpha}{(v_I^2 - v'^2)(v_I^2 - v''^2)}, \quad (28)$$

whence

$$\mathbf{D}'\mathbf{D}'' = \frac{(\mathbf{nE}')(\mathbf{nE}'')}{v'^2 - v''^2} c^4 \left\{ \frac{\cos^2 \alpha}{v_I^2 - v'^2} - \frac{\cos^2 \alpha}{v_I^2 - v''^2} + \frac{\cos^2 \beta}{v_{II}^2 - v'^2} - \frac{\cos^2 \beta}{v_{II}^2 - v''^2} + \frac{\cos^2 \gamma}{v_{III}^2 - v'^2} - \frac{\cos^2 \gamma}{v_{III}^2 - v''^2} \right\} \quad (27')$$

But since  $v'$  and  $v''$  are the roots of equation (25'), the sums of the positive and negative terms vanish separately which proves our statement.

The direction of  $\mathbf{D}$  may be determined geometrically, using the index ellipsoid: Cut the ellipsoid by a plane normal to  $\mathbf{n}$ . The principal axes of the resulting elliptical section give the directions of  $\mathbf{D}'$  and  $\mathbf{D}''$ . Proof: The principal axes are characterized by the fact that  $r^2$  is an extremum for these directions, with the auxiliary condition that the terminus of  $r$  is in the plane  $r\mathbf{n} = 0$  as well as on the ellipsoid  $r\Phi^{-1}r = 1$ . The extremal condition gives

$$r\delta r = 0, \quad (29)$$

and the variations of the auxiliary conditions are

$$\mathbf{n}\delta r = 0 \quad \text{and} \quad \delta r\Phi^{-1}r = 0^*. \quad (30)$$

Introducing the Lagrangian multipliers  $\lambda$  and  $\mu$ , we obtain the equation

$$r + \lambda\mathbf{n} + \mu\Phi^{-1}r = 0. \quad (31)$$

Scalar multiplication by  $r$  and also by  $\mathbf{n}$  yields the following equations, since  $r\mathbf{n} = 0$  and  $r\Phi^{-1}r = 1$ :

$$\begin{aligned} \mu &= -r^2, \\ \lambda &= r^2\mathbf{n}\Phi^{-1}r, \end{aligned}$$

$$\text{whence} \quad \frac{r}{r^2} + \mathbf{n} \cdot \mathbf{n}\Phi^{-1}r - \Phi^{-1}r = 0. \quad (32)$$

Since  $r = \mathbf{D}/p$  in the index ellipsoid— $p$  being a factor of proportionality—we obtain †

$$\frac{p^2}{D^2} \mathbf{D} - \Phi^{-1} \mathbf{D} + \mathbf{n} \cdot \mathbf{n}\Phi^{-1} \mathbf{D} = \frac{p^2}{D^2} \mathbf{D} - \mathbf{E} + \mathbf{n} \cdot \mathbf{nE} = 0. \quad (33)$$

\* For according to the way the index ellipsoid was introduced,  $\mathbf{E} = \Phi^{-1}(r)$  has the direction of the normal, and the condition that the variation  $\delta r$  remains on the ellipsoid means that  $\delta r$  must be perpendicular to  $\mathbf{E}$ .

† The introduction of the factor  $p$  is necessary when comparing equation (33) with

Identifying  $p^2/D^2$  with  $v^2/c^2$ , we see that this is precisely equation (22) (p. 364). Hence the construction is justified.

Having thus found the two vectors  $\mathbf{D}'$  and  $\mathbf{D}''$  corresponding to a normal  $\mathbf{n}$ , the directions of the associated vectors  $\mathbf{E}'$  and  $\mathbf{E}''$  are given as those of the normals to the ellipsoid through the termini of  $\mathbf{D}'$  and  $\mathbf{D}''$ . These are, in general, *not* mutually perpendicular. The vectors  $\mathbf{H}'$  and  $\mathbf{H}''$  are normal to the planes of  $\mathbf{D}'$  and  $\mathbf{n}$  and  $\mathbf{D}''$  and  $\mathbf{n}$  respectively. There are two rays corresponding to each wave-normal, viz.:

$$\mathbf{S}' = \frac{c}{4\pi} [\mathbf{E}'\mathbf{H}'] \quad \text{and} \quad \mathbf{S}'' = \frac{c}{4\pi} [\mathbf{E}''\mathbf{H}''].$$

We designate the corresponding unit vectors by  $\mathbf{s}'$  and  $\mathbf{s}''$ .

While the direction of  $\mathbf{n}$  leaves an ambiguity, that of  $\mathbf{D}$  determines  $\mathbf{n}$  as well as  $\mathbf{E}$  and  $\mathbf{S}$ , and the magnitude of  $v$  uniquely; for  $\mathbf{n}$ , on one hand, must be in the plane normal to  $\mathbf{D}$ , and, in addition, the vector  $\mathbf{E}$  is uniquely determined for a given  $\mathbf{D}$  by the index ellipsoid. Since  $\mathbf{n}$  must lie also in the plane of  $\mathbf{D}$  and  $\mathbf{E}$ , the intersection of the two planes gives the direction of  $\mathbf{n}$ . The magnitude of  $v$  is obtained most simply from equation (22) (p. 364): If we multiply this equation scalarly by  $\mathbf{D}$ , we have

$$v = \frac{c\sqrt{\mathbf{E}\mathbf{D}}}{\sqrt{\mathbf{D}^2}} = c \frac{\sqrt{\mathbf{D}\Phi^{-1}\mathbf{D}}}{\sqrt{\mathbf{D}^2}}. \quad \dots \dots (34)$$

On the other hand, if we start with the direction  $\mathbf{s}$  of the ray, we obtain completely analogous relationships in which  $\mathbf{D}$  and  $\mathbf{E}$  have interchanged their rôles. The fact that  $\mathbf{D}$ ,  $\mathbf{E}$  and  $\mathbf{s}$  lie in one plane may be expressed by the equation (cf. *Ex. 2*, p. 10):

$$\mathbf{D} + \alpha\mathbf{E} + \beta\mathbf{s} = 0. \quad \dots \dots (35)$$

The coefficients  $\alpha$  and  $\beta$  are obtained by scalar multiplication of this equation by  $\mathbf{s}$  and by  $\mathbf{n}$  in turn: Since  $\mathbf{s}$  is normal to  $\mathbf{E}$ , multiplication by  $\mathbf{s}$  gives

$$\beta = -\mathbf{D}\mathbf{s},$$

and since  $\mathbf{n}$  is normal to  $\mathbf{D}$ , multiplication by  $\mathbf{n}$  yields, after inserting the value of  $\beta$ ,

$$\alpha = \frac{(\mathbf{D}\mathbf{s})(\mathbf{s}\mathbf{n})}{\mathbf{E}\mathbf{n}}.$$

But from equation (22) (p. 364) we have

$$\mathbf{D}\mathbf{s} = -\frac{c^2}{v^2}(\mathbf{n}\mathbf{s})(\mathbf{E}\mathbf{n}),$$

equation (22), for otherwise if we put  $\mathbf{r} = \mathbf{D}$ , comparison of these equations yields the meaningless result  $D^2 = c^2/v^2$ .



so that

$$a = -\frac{c^2}{v^2} (ns)^2 = -\frac{c^2}{v^2} \cos^2 \theta = -\frac{c^2}{V^2}.$$

Then we have

$$\frac{c^2}{V^2} \mathbf{E} - \mathbf{D} + \mathbf{s} \cdot \mathbf{s} \mathbf{D} = 0. \quad . \quad . \quad . \quad . \quad (36)$$

Comparison of this equation with (22) shows that  $\mathbf{D}$  and  $\mathbf{E}$  are interchanged, and that  $c^2/V^2$  appears in place of  $v^2/c^2$ .

If we denote the components of  $\mathbf{s}$ —i.e. the direction cosines of the ray—by  $\cos A$ ,  $\cos B$ ,  $\cos \Gamma$ , and express the first  $\mathbf{D}$  in (36) in terms of  $\mathbf{E}$ , we obtain the component equations

$$E_x = \frac{\mathbf{D} \mathbf{s} \cos A}{K_I - \frac{c^2}{V^2}}, \quad E_y = \frac{\mathbf{D} \mathbf{s} \cos B}{K_{II} - \frac{c^2}{V^2}}, \quad E_z = \frac{\mathbf{D} \mathbf{s} \cos \Gamma}{K_{III} - \frac{c^2}{V^2}}. \quad (37)$$

Using these components to express the vanishing of the scalar product  $\mathbf{E} \mathbf{s}$  and replacing  $K_I, \dots$  by  $c^2/v_I^2, \dots$ , we obtain the equation

$$\frac{v_I^2 \cos^2 A}{v_I^2 - V^2} + \frac{v_{II}^2 \cos^2 B}{v_{II}^2 - V^2} + \frac{v_{III}^2 \cos^2 \Gamma}{v_{III}^2 - V^2} = 0. \quad . \quad . \quad (38)$$

This is a quadratic equation for  $V^2$  as a function of the direction of the ray. The directions of the corresponding  $\mathbf{E}'$  and  $\mathbf{E}''$  are obtained as for  $\mathbf{D}'$  and  $\mathbf{D}''$  above, either analytically from equation (37) or geometrically as the axes of the ellipse cut from the Fresnel ellipsoid by a plane through the centre and normal to  $\mathbf{s}$ . The corresponding vectors  $\mathbf{D}'$  and  $\mathbf{D}''$ , which are not mutually perpendicular, are then found from the ellipsoid in the usual way. As demonstrated for  $\mathbf{D}$  above, it also may be shown that  $\mathbf{E}$  uniquely determines the wave.

### 3. Normal Surface and Wave Surface. The Optical Axes.

We now consider an optical disturbance spreading outwards from a point. This point need not be an actual *source* of light, but, as we shall see in the next chapter, may be any point of a wave front. We shall find that under certain conditions the so-called Principle of Huygens permits us to consider any such point as a new source of light. We may represent the light emanating from  $O$  in such way that a surface element—a portion of a plane wave front—travels in each direction. After one second, these wave fronts have travelled a distance  $v$  in the direction of their normals. If we measure off the corresponding segments  $OP = r = v$  on these normals, the end points form a surface called the *normal surface*. However, each of the actual elementary wave fronts does not lie on this surface, but somewhere

to the side of the corresponding point  $P$  (see fig. 2). Moreover, all elementary wave fronts emanating from  $O$  lie on a new surface called the *ray surface* or *wave surface*. First let us consider the normal surface. We obtain its equation by putting  $r$  in place of  $v$  and  $x/r, \dots$  in place of  $\cos \alpha, \dots$  in equation (25') (p. 363):

$$(v_{\text{II}}^2 - r^2)(v_{\text{III}}^2 - r^2)x^2 + (v_{\text{III}}^2 - r^2)(v_{\text{I}}^2 - r^2)y^2 + (v_{\text{I}}^2 - r^2)(v_{\text{II}}^2 - r^2)z^2 = 0. \quad (39)$$

This is a surface of the sixth degree and consists of two sheets, corresponding to the fact that there are two phase velocities for every direction. We wish to consider the curves of intersection of this surface with the co-ordinate planes, which—as planes of symmetry of the tensor ellipsoid—are of particular importance. The intersection with the plane  $y = 0$ , for example, yields

$$(v_{\text{II}}^2 - r^2)(v_{\text{III}}^2 - r^2)x^2 + (v_{\text{I}}^2 - r^2)(v_{\text{II}}^2 - r^2)z^2 = 0. \quad (39')$$

This curve consists of the circle  $r^2 = v_{\text{II}}^2$  and the oval of fourth degree  $(v_{\text{III}}^2 - r^2)x^2 + (v_{\text{I}}^2 - r^2)z^2 = 0$ . If

we now assume that  $v_{\text{I}} < v_{\text{II}} < v_{\text{III}}$  we see that the circle and the oval have real points of intersection in the plane  $y = 0$ , and only in this plane; for if we put  $r = v_{\text{II}}$  in the equation of the oval, the coefficients of  $x^2$  and  $z^2$  have different signs, and so the equation can be satisfied by real values of  $x$  and  $z$ . The directions of the lines drawn from the centre to these points of intersection are given by

$$\frac{z}{x} = \pm \sqrt{\frac{v_{\text{III}}^2 - v_{\text{II}}^2}{v_{\text{II}}^2 - v_{\text{I}}^2}}. \quad (40)$$

These directions are called the *optical axes*, since in these directions there is but one phase velocity  $v$ . The wave normals coincide with one of the axes, and we take the one that goes with the positive sign of the root. The direction cosines of the normals are

$$\left. \begin{aligned} \cos \alpha &= \frac{1}{\sqrt{(1 + \tan^2 \alpha)}} = \sqrt{\frac{v_{\text{III}}^2 - v_{\text{I}}^2}{v_{\text{III}}^2 - v_{\text{II}}^2}}, & \cos \beta &= 0, \\ \cos \gamma &= \sin \alpha = \sqrt{\frac{v_{\text{III}}^2 - v_{\text{II}}^2}{v_{\text{III}}^2 - v_{\text{I}}^2}}. \end{aligned} \right\}. \quad (41)$$

Setting these values into (25') we get the single value of the phase velocity  $v' = v'' = v_{\text{II}}$ . But then the  $Y$ -component of  $\mathbf{D}$  is indeter-

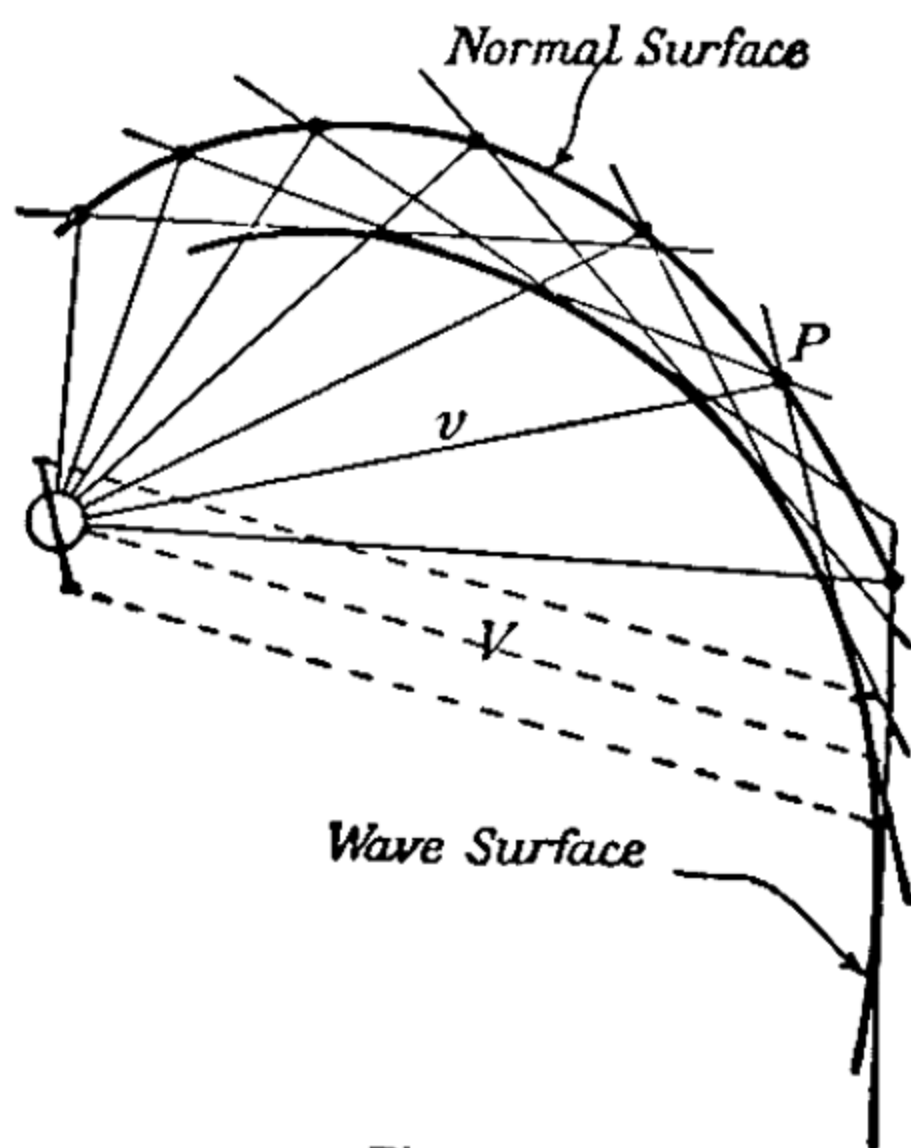


Fig. 2



minate, since  $\cos \beta$  as well as  $(1/K_{II}) - (v_{II}^2/c^2)$  vanish in (24). For  $D_z/D_x$  we get, as we must,  $-\cot \alpha$ , since  $\mathbf{D}$  is always normal to  $\mathbf{n}$ . Whatever value we assign to the  $Y$ -component,  $\mathbf{D}$  remains perpendicular to  $\mathbf{n}$ , and so  $\mathbf{D}$  may assume any direction in the plane normal to  $\mathbf{n}$ .

On the other hand, the geometric construction must always furnish *two* axes of the ellipse of intersection. The contradiction is only an apparent one, the explanation being that the elliptical sections are circles in this case. Hence *the optical axes are normal to those sections of the index ellipsoid which are circles*.

Since the axial direction lies in the  $xz$ -plane, the  $y$ -axis is one of the diameters of these circular sections. This is important in what follows. But since a circle has no uniquely determined axes, the direction of  $\mathbf{D}$  is undetermined. Thus the associated ray is also undetermined. It is not correct to conclude, however, that the various possible directions for the rays are obtained by rotating the entire vector system of fig. 1 (p. 364) about  $\mathbf{n}$ , for the angle between  $\mathbf{D}$  and  $\mathbf{E}$ —which is the same as that between  $\mathbf{n}$  and  $\mathbf{s}$ —depends entirely upon the position of  $\mathbf{D}$ . Moreover,  $\mathbf{n}$  is not the axis of the cone containing the several rays, but is itself one of the generators; since the  $y$ -axis is a diameter of a circular section, it is also a permissible direction for  $\mathbf{D}$ , but  $\mathbf{D}$  and  $\mathbf{E}$  have the same direction in this instance, so that  $\mathbf{s}$  and  $\mathbf{n}$  coincide.

If the magnitude of  $v_{III}$  approaches that of  $v_{II}$ , the angle between the two axes becomes smaller and smaller; in the limit  $v_{III} = v_{II}$ , both axes coincide with the  $x$ -axis, and the circle touches the oval at the end points of the axes which coincide. This is the case for a so-called *optically uniaxial crystal*. Here the Fresnel ellipsoid and the index ellipsoid are figures of rotation and, as seen from equation (39) above, the normal surface degenerates into the sphere  $r^2 = v_{II}^2$  and the fourth-order surface of rotation

$$(v_{II}^2 - r^2)x^2 + (v_I^2 - r^2)(y^2 + z^2) = 0. \quad . \quad . \quad (39'')$$

Of greater importance than the normal surface is the *ray surface* or *wave surface*, obtained by laying off the ray velocity along the direction of the ray. The elements of the actual wave front are at these points, and are given by the tangent planes to this surface. The feet of perpendiculars drawn to these planes from the origin lie in the normal surface, which is therefore a so-called *pedal-surface* of the wave surface. The equation of the wave surface is obtained by putting  $r$  for  $V$  and  $x/r \dots$  for  $\cos A \dots$  in equation (38) (p. 368):

$$v_I^2(v_{II}^2 - r^2)(v_{III}^2 - r^2)x^2 + v_{II}^2(v_{III}^2 - r^2)(v_I^2 - r^2)y^2 + v_{III}^2(v_I^2 - r^2)(v_{II}^2 - r^2)z^2 = 0. \quad . \quad . \quad (42)$$

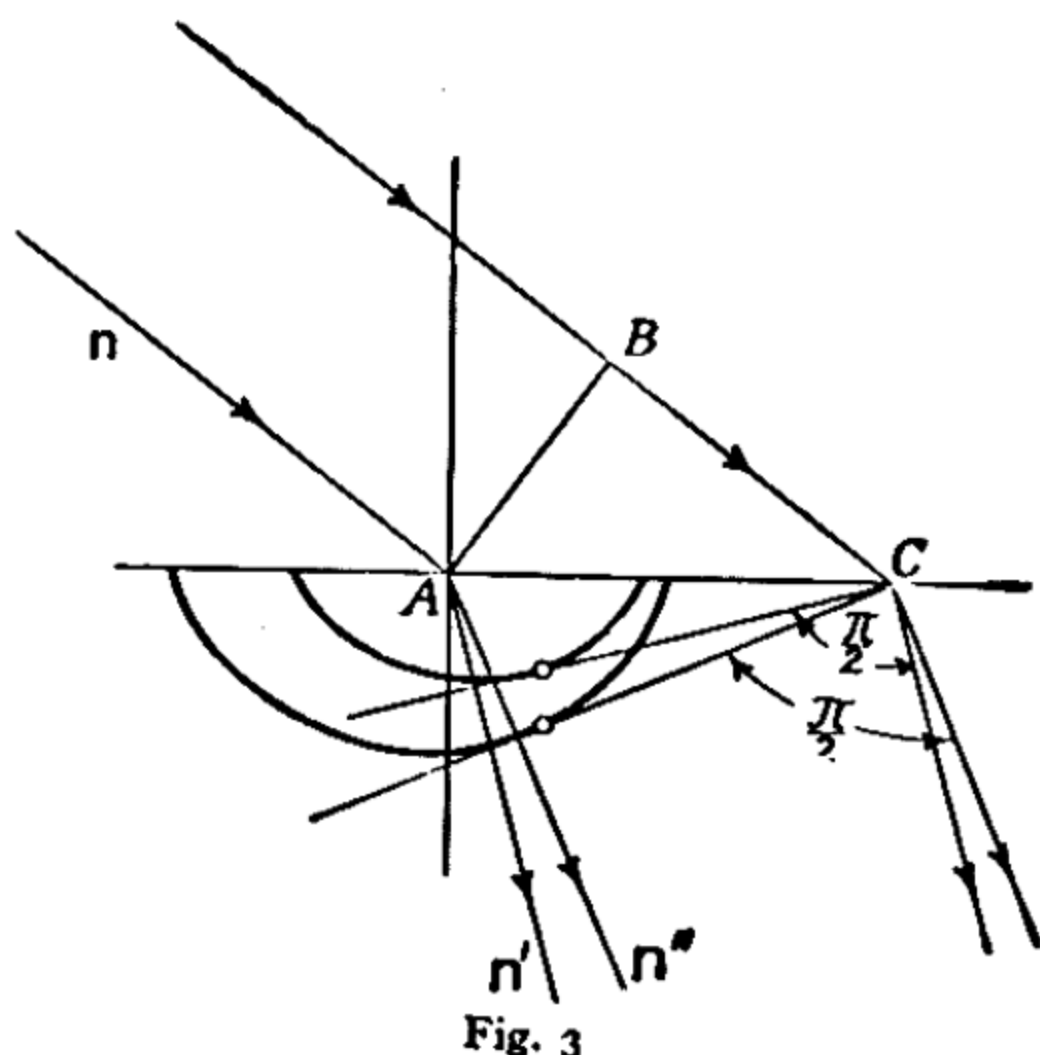


The term of lowest degree is  $v_I^2 v_{II}^2 v_{III}^2 (x^2 + y^2 + z^2)$ . Since all other terms contain  $r^2$ , we may divide throughout by  $r^2$ , and the degree of the ray surface is not six but four. Correspondingly, the curves of intersection with the co-ordinate planes are circles and ellipses (cf. fig. 3, Plate I, facing p. 240). There is again a plane in which the circle and ellipse have real points of intersection. The lines joining the opposite pairs of these points are called the optical *ray axes*. If the ray has the direction of one of the optical ray axes, there are infinitely many normal directions corresponding to one ray, and these normals form the surface of a cone (see below). For uniaxial crystals the wave surface degenerates into the sphere  $r^2 = v_{II}^2$  and the prolate ellipsoid of rotation

$$\frac{x^2}{v_{II}^2} + \frac{y^2 + z^2}{v_I^2} - 1 = 0. \quad \checkmark \quad \dots \quad (43)$$

#### 4. Refraction of Plane Waves at the Plane Surface of an Anisotropic Medium.

Since the continuity of the tangential components of  $\mathbf{E}$  and of  $\mathbf{H}$  (p. 345) must hold for anisotropic media also, we can make use of the considerations of § 3, Chap. XVIII (p. 349), from which we obtain the laws of reflection and refraction. It must be remembered, however, that  $\mathbf{n}$  denotes the direction of the wave normal, as may be seen from equation (16) (p. 363). Hence, at the boundary of an anisotropic body the *wave normal* remains in the plane of incidence and obeys the Law of Refraction. But the recognition of this fact is not of much aid in constructing the refracted wave normal, for, in contrast with isotropic media, the value of the index of refraction now depends upon the direction of this normal. This means that the direction of the refracted normal must be known in advance, in order to find the correct value of the refractive index. Fortunately, there is available a method which leads to the desired result. This method is often used to furnish a simple derivation of Snell's law for isotropic media, and is based upon Huygens' Principle. According to this principle, which will be discussed in detail in the next chapter, each point on a wave front may be taken to be the source of secondary wavelets: the actual wave



front at any future time may be considered to be the envelope of the secondary waves. Let us apply this idea to a plane wave incident upon a plane surface of a crystal (fig. 3). We must then consider the secondary waves emanating from points in the surface, at the same time remembering that the arrival of the primary disturbance takes place at different times, since points in the same phase are on the line  $AB$ . By the time the incident wave has arrived at  $C$ , the disturbance coming from  $A$  has already spread out over a wave surface. Assuming the upper medium to be a vacuum, this corresponds to a time  $BC/c$ . The wave surfaces belonging to points lying between  $A$  and  $C$  should be drawn in corresponding ratio, but since the envelope of these wave surfaces is a plane passing through  $C$ , it is sufficient to take the plane through  $C$  tangent to the wave surface of  $A$  as the new wave front. Since the wave front is of indefinite extent in a direction normal to the plane of the figure, we really mean the plane containing the line through  $C$  normal to the diagram and tangent to the family of wave surfaces originating in points on a line through  $A$  normal to the figure:

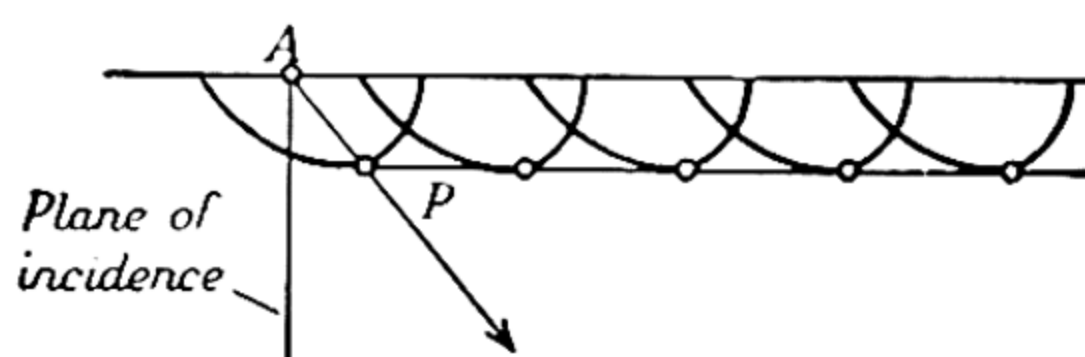


Fig. 4

Imagine a ray determined by a small surface element in the neighbourhood of  $A$ . After refraction, this element of the wave front must contain the point of tangency of the wave front with the wave surface corresponding to the point  $A$ . What happens for one wave surface in a plane normal to that of fig. 3 is shown in fig. 4. The top line in the latter diagram projects as the point  $A$  in the former figure. The two points of contact—one corresponding to each sheet—do not, in general, lie in the plane of fig. 4. Hence a *refracted ray does not, in general, follow the law of refraction.*

According to what we have just seen, it is important to differentiate between refraction of the normal and refraction of a ray, i.e. a limited portion of a wave front. This difference becomes evident if we consider the simplest instance of the passage of light through a plane parallel plate of anisotropic material. If a plane wave is incident normally upon such a plate, the normal will not be deviated; nevertheless, there are two wave planes within the crystal perpendicular to the normal direction. These correspond to the two velocities of propagation belonging to each direction. If a plane parallel crystal plate is inserted in the parallel bundle of rays between the collimator



and telescope of a spectroscope, no double refraction is observed. The effect is quite different, however, if a narrow bundle of rays is used in place of a wide beam; in this instance double refraction takes place, for to each normal direction there correspond, in general, two rays. The results are particularly remarkable when the wave normal is in the direction of one of the optic axes of a biaxial crystal. In this case there are infinitely many rays corresponding to the normal, for it is possible to draw a tangent plane to the two sheets of the ray surface such that the plane touches the surface on a circle (a portion of this circle is seen in fig. 3, Plate I). Hence any ray from the centre which passes through this circle corresponds to this plane wave front, and the transmitted rays form the surface of a cone, which in turn is refracted into the form of a hollow cylinder of light upon emerging. This phenomenon is known as *internal conical refraction*.

The effect reciprocal to double refraction of a ray is obtained by allowing light to enter a biaxial crystal in all directions through a given point on the surface of the plate. This may be done by diffuse illumination of a pinhole in a piece of cardboard placed on the plate. A given ray may be isolated by placing a similar diaphragm on the opposite side of the plate. This ray, however, will experience double refraction on emerging, since there are two different corresponding normal directions, each having its own value of the refractive index. If the transmitted ray is allowed to have the direction of one of the optical ray axes, there will be an infinity of normals corresponding to this so-called *single-ray direction*. This ray is directed from the centre to the "dimple" in the wave surface (fig. 3, Plate I), at which point it is possible to draw an infinite number of planes tangent to the surface. The emerging rays fill a cone, and this phenomenon is known as *external conical refraction*. These conical refraction effects were predicted by Hamilton and were verified by Lloyd, using a crystal of aragonite.

In the previous discussion we tacitly assumed the incident light to be unpolarized, i.e. there was no uniquely characterized plane of vibration of  $\mathbf{E}$ . Let us return to a consideration of the normal incidence of a plane wave upon a plane parallel plate. There are two mutually perpendicular directions of  $\mathbf{D}$ ; since these are both normal to  $\mathbf{n}$  they lie in the surface of the plate. As before, we denote these two vectors by  $\mathbf{D}'$  and  $\mathbf{D}''$ .  $\mathbf{D}$  and  $\mathbf{n}$  also determine the plane of vibration of  $\mathbf{H}$  (cf. fig. 1, p. 364). If, now, the incident light is linearly polarized, with its magnetic vector forming an angle  $\alpha$  with one of the vectors  $\mathbf{D}$ , then we can resolve the magnetic vector into two components, one along  $\mathbf{D}'$  and one along  $\mathbf{D}''$ . But each of the components has its own phase velocity, corresponding to the two different velocities belonging to  $\mathbf{n}$ . As a result, there is a difference in phase when the light emerges—the original linearly



polarized light has become elliptically polarized. On the other hand, by introducing a phase difference in this manner, we can convert elliptically polarized light into plane polarized light. This is the principle of so-called compensators for polarized light.

Finally, we wish to consider briefly the beautiful patterns shown by plates of anisotropic materials when placed between crossed Nicol prisms. If we allow polarized light to fall upon a crystal plate at angle of incidence  $\phi$ , then within the crystal there are two distinct normal directions, each with one plane of vibration of  $\mathbf{D}$ , and hence of  $\mathbf{E}$ .\* In general, the plane of vibration of the incident light will not coincide with either of these directions, and so there will be a component

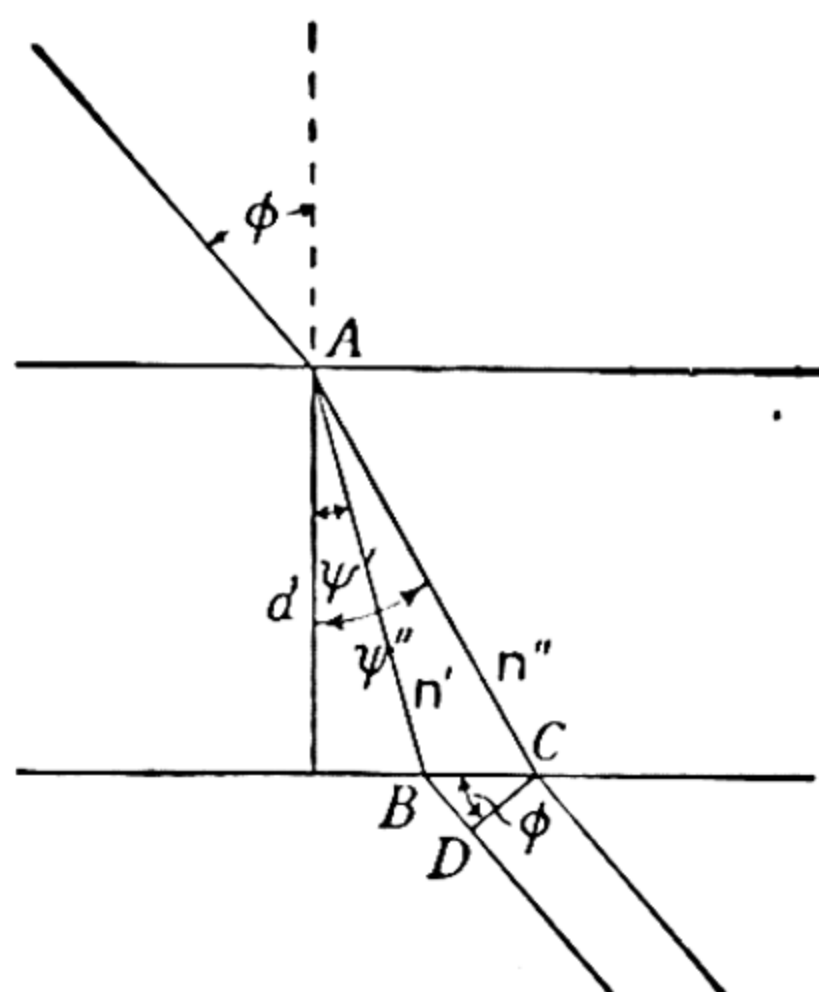


Fig. 5

along each. Both wave normals again have the same direction upon emerging, since we can apply the same construction here as we did at the point of entry into the crystal (reversibility of the rays). But since the two waves have travelled over different paths within the crystal they will differ in phase; this means that the emergent light is, in general, elliptically polarized (cf. the special case of normal incidence treated above). From fig. 5 we see that the phase difference is given by

$$\delta = \omega \left( \frac{AB}{v'} + \frac{BD}{c} - \frac{AC}{v''} \right).$$

Denoting the angles between the normal to the surface and the re-

\* It would be wrong to assume that there are two planes of vibration corresponding to each of the refracted normals. As a matter of fact, the construction of the refracted wave front shows that corresponding to each of the two normals there is a definite normal velocity  $v$ , and hence by equation (26) (p. 365), a definite direction of  $\mathbf{D}$ .

fracted wave normals by  $\psi'$  and  $\psi''$  respectively, and calling the thickness of the plate  $d$ , we have

$$AB = \frac{d}{\cos \psi''}; \quad AC = \frac{d}{\cos \psi'}; \quad BD = BC \sin \phi = d (\tan \psi'' - \tan \psi') \sin \phi,$$

i.e.

$$\delta = \omega d \left\{ \left( \frac{\sin \psi'' \sin \phi}{c} - \frac{1}{v''} \right) \frac{1}{\cos \psi''} - \left( \frac{\sin \psi' \sin \phi}{c} - \frac{1}{v'} \right) \frac{1}{\cos \psi'} \right\},$$

Since the law of refraction holds for the wave normal,

$$\frac{\sin \phi}{c} = \frac{\sin \psi''}{v''} = \frac{\sin \psi'}{v'},$$

or

$$\delta = \omega d \left\{ \frac{\cos \psi'}{v'} - \frac{\cos \psi''}{v''} \right\}, \quad . . . . (44)$$

and for the common case of approximately normal incidence,

$$\delta \approx \omega d \left( \frac{1}{v'} - \frac{1}{v''} \right). \quad . . . . (44')$$

If the phase difference is  $2n\pi$  the emergent wave is again linearly polarized in the original plane, and can be extinguished by a second Nicol prism turned at an angle of  $\pi/2$  with respect to the polarizer. If we allow convergent polarized light to fall upon the crystal, there will be darkness for all directions  $\phi$  for which the angles of refraction  $\psi'$  and  $\psi''$  yield  $\delta = 2n\pi$  according to (44). These directions then give dark curves in the focal plane of a lens. But these are not the only dark curves. There is a second possibility: for a given angle of incidence the plane of vibration of the incident wave can coincide with one of the possible planes of vibration in the crystal. In this case there is no second wave to bring about elliptic polarization. The resulting dark curve is called an *isogyre*.

*Ex. 98.* A plate of a uniaxial crystal is cut perpendicular to the axis. Show that with crossed Nicols a system of circular rings can be obtained, with a rectangular cross through their centre.

## CHAPTER XX

### ELECTROMAGNETIC WAVES. IV: THE THEORY OF DIFFRACTION

#### ✓ 1. The General Diffraction Problem and Attempts to solve it. Kirchhoff's Formula.

In previous sections we have dealt almost always with wave fronts of unlimited extent. In those cases where it proved necessary to consider limited beams, in the study of crystal optics, for example, we assumed that the limited beam was propagated in the same way as an extended one, and that the cross-section of the beam remained as determined by the aperture through which the light passed. Actually, this is not strictly true. 'Simple experiments show that light enters that portion of space which, according to our previous assumptions, could not be reached by the light wave—the region of the so-called *geometric shadow*. This phenomenon is called *diffraction*.'

'Mathematically, the problem of the diffraction of a wave by an obstacle is to be formulated in this way: we seek a solution of the wave equation for  $\mathbf{E}$  and  $\mathbf{H}$  which satisfies the differential equation both for the region within the diffracting body, or obstacle, and for the rest of space; and which, moreover, satisfies the condition of continuity of the tangential components of  $\mathbf{E}$  and  $\mathbf{H}$  at the bounding surface. At the same time it must be remembered that the electrical constants ( $K, \mu, \sigma$ ) are different for the two regions.<sup>1</sup> If it were possible to find such a solution, it would be expected to furnish information on all questions of intensity, polarization, &c., behind the obstacle. However, the mathematical difficulties connected with such a solution are so great that it has been possible, so far, to develop solutions only in especially simple cases, such as the diffraction of a plane wave meeting a sphere or a cylinder or a screen.

'For the greater number of practically important diffraction problems we have to fall back upon an approximate method, which proves adequate if the conditions are not too unfavourable. Essentially, this method is merely the mathematical formulation of the principle that every point of a wave may be considered to be the source of secondary spherical waves. This concept we have already utilized in the preceding chapter. According to the original Huygens formulation, the wave front existing after a certain time is given by the



surface which envelops the elementary wavelets. We made use of the principle in this simple form in our discussion of the refraction of light by crystals (p. 371). Fresnel, however, improved the method in this way: in order to find the resultant disturbance at a point of observation  $P$ , he imagined the secondary waves arriving there at a certain instant to be superimposed, at the same time making allowance for their difference in phase. The mathematical statement of this idea is contained in a formula due to Kirchhoff, which we shall now prove.

(We represent the light disturbance by a scalar quantity  $U$ ; this may be interpreted as one of the two components of  $\mathbf{E}$ , which are perpendicular to the wave normal. The only essential fact is that the intensity of the light is proportional to  $U^2$ .  $U$  must satisfy the wave equation

$$\Delta U = \frac{n^2}{c^2} \frac{\partial^2 U}{\partial t^2}. \quad . . . . . (1)$$

For the dependence upon time we make the usual assumption for a monochromatic vibration

$$U = u(x, y, z)e^{i\omega t}, \quad . . . . . (2)$$

and obtain the following differential equation for the spatial part, which alone interests us:

$$\Delta u + k^2 u = 0. \quad . . . . . (3)$$

Here

$$k = \frac{n\omega}{c} = \frac{2\pi}{\lambda}. \quad . . . . . (4)$$

Assume now that we have two functions,  $u$  and  $v$ , which satisfy the wave equation. We then apply Gauss's theorem to the vector quantity  $u \text{ grad } v - v \text{ grad } u$ , integrating over an arbitrary volume in which the integrand is continuous. The result is Green's theorem (p. 270):

$$\begin{aligned} \oint (u \text{ grad } v - v \text{ grad } u) d\mathbf{S} &= \int (u \Delta v - v \Delta u) d\tau \\ &= \int k^2 (vu - uv) d\tau = 0. \end{aligned} \quad (5)$$

We take the function  $v$  to be the spherical wave originating at a point  $P$  in the region of integration. Apart from the time factor, this is

$$v = \frac{e^{-ikr}}{r}. \quad . . . . . (6)$$

But this function becomes infinite at  $P$ ; hence if we wish to apply equation (5) we must surround this point by a small sphere. The surface of the sphere is then part of the bounding surface of the region of integration, and the point  $P$  is thus excluded from this space. The

normal is always directed outward from the region of integration; hence in this case the normal is directed toward the centre of the small sphere (fig. 1). We now have the equation

$$\oint_s \left( u \operatorname{grad} \frac{e^{-ikr}}{r} - \frac{e^{-ikr}}{r} \operatorname{grad} u \right) d\mathbf{S} + \oint_{\text{Sphere}} \left( u \operatorname{grad} \frac{e^{-ikr}}{r} - \frac{e^{-ikr}}{r} \operatorname{grad} u \right) d\mathbf{S} = 0. \quad (7)$$

The integral over the surface of the small sphere may be calculated readily. First we note that since  $u$  and  $\operatorname{grad} u$  are continuous, the contribution of the term

$$\frac{e^{-ikr}}{r} \operatorname{grad} u$$

vanishes when we allow the sphere to become infinitely small, since  $d\mathbf{S}$  decreases as  $r^2$ , while we have only the first power of  $r$  in the denominator. Again, since  $u$  is continuous, we may replace it by the value  $u_P$  at the centre when integrating over the small sphere. Finally, we have

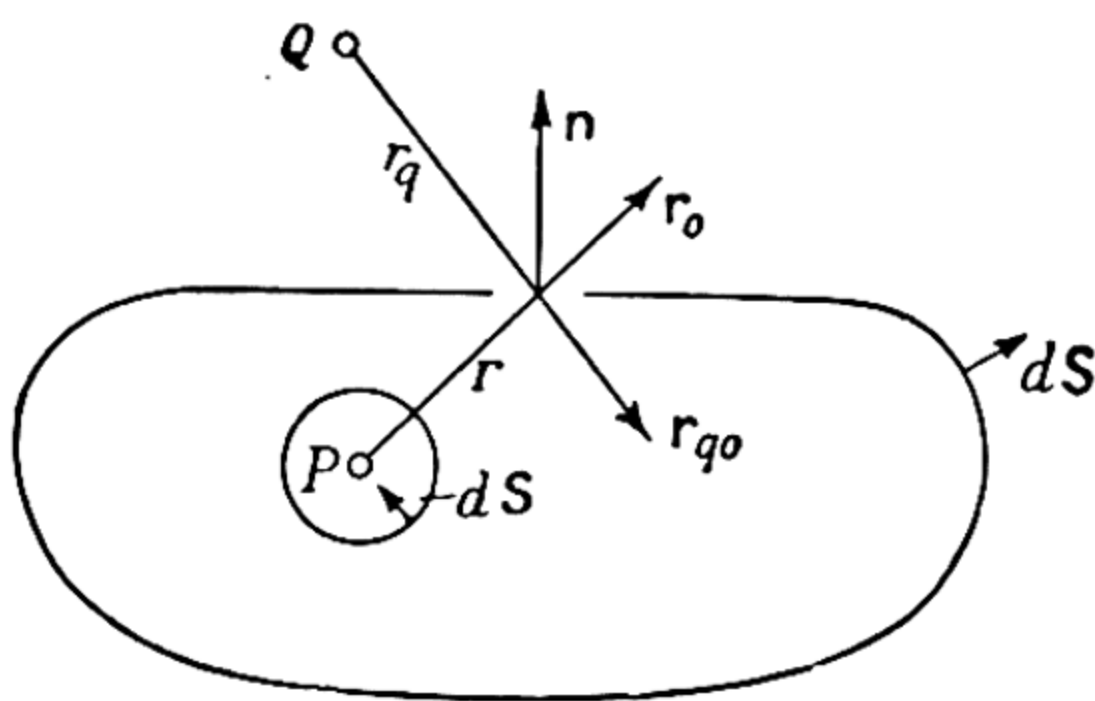


Fig. 1

$$\operatorname{grad} \frac{e^{-ikr}}{r} = - \left( \frac{1}{r^2} + \frac{ik}{r} \right) e^{-ikr} \mathbf{r}_0, \quad \dots (8)$$

where  $\mathbf{r}_0$  is a unit vector directed outwards from the centre of the sphere. Hence we have

$$d\mathbf{S} = -\mathbf{r}_0 dS = -\mathbf{r}_0 r^2 \sin \theta d\theta d\phi, \quad \dots (9)$$

and the value of the integral

$$u_P \oint_{\text{Sphere}} \operatorname{grad} \frac{e^{-ikr}}{r} d\mathbf{S}$$

in the limit ( $r = 0$ ) becomes  $4\pi u_P$ . Hence

$$u_P = \frac{1}{4\pi} \oint_s \left( \frac{e^{-ikr}}{r} \operatorname{grad} u - u \operatorname{grad} \frac{e^{-ikr}}{r} \right) d\mathbf{S}. \quad (10)$$

This is the celebrated formula of Kirchhoff.\* It expresses the disturbance at a point  $P$  as the integral of the light disturbance over the boundary of a region containing  $P$ . If, for example, we are dealing with the

\*The formula was first proved by Helmholtz, but is a special case of a formula due to Kirchhoff.

diffraction of light by a small aperture in a screen, the screen would be taken as the boundary of the region. But in order to apply the formula, the diffraction problem must, in a way, be solved already, for one must know the disturbance in the opening and on the surface of the screen in advance. Kirchhoff made the natural assumption—which is amply justified by experience—that the disturbance in the free aperture is that which would exist if the screen were not present, and that  $u$  and  $\text{grad } u$  are zero on the screen itself. It is evident that with an assumption of this kind, which contains nothing concerning the material properties of the screen, we cannot attack the more delicate questions, such as that of polarization. Nevertheless, if we do not require the distribution of light in the immediate neighbourhood of the edge, the Kirchhoff assumptions prove adequate.

## 2. Reciprocal Theorems of the Theory of Diffraction. Classification of Diffraction Phenomena.

We shall now apply the Kirchhoff formula to the special case where the primary excitation is from a point source  $Q$  and is represented by

$$u = \frac{e^{-ikr_q}}{r_q} \quad \dots \dots \dots (11)$$

Then we have

$$\left. \begin{aligned} \text{grad } u &= - \left( \frac{1}{r_q^2} e^{-ikr_q} + \frac{ik}{r_q} e^{-ikr_q} \right) \mathbf{r}_{q0}, \\ \text{grad } \frac{e^{-ikr}}{r} &= - \left( \frac{1}{r^2} e^{-ikr} + \frac{ik}{r} e^{-ikr} \right) \mathbf{r}_0 \end{aligned} \right\}, \quad \dots \quad (12)$$

where  $\mathbf{r}_q$  is the radius vector drawn from  $Q$  to a point of the aperture and  $\mathbf{r}$  is the radius vector drawn from the point of observation  $P$ . The corresponding unit vectors are designated by the subscript 0 (fig. 1). Since  $\mathbf{r}_{q0} d\mathbf{S} = \cos(\mathbf{n}\mathbf{r}_q) dS$ , we have

$$u_P = \frac{1}{4\pi} \int_{\text{Aperture}} \frac{e^{-ik(r+r_q)}}{rr_q} \left\{ \left( \frac{1}{r} + ik \right) \cos(\mathbf{n}\mathbf{r}) - \left( \frac{1}{r_q} + ik \right) \cos(\mathbf{n}\mathbf{r}_q) \right\} dS. \quad (13)$$

We now assume that the distances from the source to the aperture and from the point of observation to the aperture are large compared with the wave-length. This condition is fulfilled in all cases occurring in practice. We may therefore neglect  $1/r$  and  $1/r_q$  when compared with  $k = 2\pi/\lambda$ , and so obtain

$$u_P = \frac{i}{2\lambda} \int \frac{e^{-ik(r+r_q)}}{rr_q} \{ \cos(\mathbf{n}\mathbf{r}) - \cos(\mathbf{n}\mathbf{r}_q) \} dS. \quad (14)$$



Assume further that both  $r$  and  $r_0$  are large compared with the dimensions of the aperture—a condition which is also realized in all practical cases. This permits us to look upon  $r$  and  $r_0$  as constants where they occur in the denominator and in the trigonometric functions of the integrand; however, this is not allowable where these quantities occur in the exponent, since they are multiplied by the large factor  $k = 2\pi/\lambda$ . Bringing the constant values outside the sign of integration, we obtain the simpler formula

$$u_P = \frac{i \{ \cos(nr) - \cos(nr_0) \}}{2\lambda r r_0} \int e^{-ik(r+r_0)} dS. \quad (15)$$

Formulae (13) to (15) are entirely symmetrical in the co-ordinates of light source and point of observation—interchanging these quantities merely alters the algebraic sign. This means that if a source  $Q$  causes a certain intensity at a point  $P$ , the identical source placed at  $P$  would cause this same intensity at the point  $Q$ . However, this must not be interpreted as a reversibility of the light path for diffraction in the sense that the use, as a source, of a distribution of light corresponding to the diffraction pattern of a point source would, in turn, have a single point of light as a diffraction pattern, i.e. the aperture would act as a lens. The theorem says nothing regarding the intensity in the neighbourhood of  $Q$ .

A second very remarkable theorem was discovered by Babinet. Imagine a number of apertures of arbitrary shape cut from an opaque screen. Let these pieces which were cut out be placed in a large opening, this system then being considered a second screen. Two such screens, for which the openings in one are the same as the opaque portions of the other, are said to be complementary. Babinet's theorem states that the diffraction patterns produced by complementary screens are identical, except for the central spot corresponding to the direct light from the source. To demonstrate this we may imagine an image of the source formed by a lens of large aperture. In the limiting case of full aperture there would be no light outside the geometric image. Hence at a point  $P$  in the neighbourhood of this image the disturbance  $u$  would be zero. We now insert one of the screens between source and image. Let this cause the disturbance  $u_1$  at  $P$ . Let the disturbance produced by the complementary screen alone be  $u_2$ . For no screen at all, the disturbance at  $P$ —which is zero—would be given by the sum of the integrals over both complementary screens. Hence by Kirchhoff's formula,

$$0 = u_1 + u_2, \\ \text{and} \quad u_1^2 = u_2^2, \quad . \quad . \quad . \quad . \quad . \quad . \quad (16)$$

which proves the theorem. //



classification of diffraction phenomena: If source and observer are so far from the screen that we may neglect the terms of  $\phi(\xi, \eta)$  which are of the second degree in  $\xi, \eta$ , we speak of *Fraunhofer diffraction phenomena*; in other cases we refer to *Fresnel diffraction phenomena*.

If we insert the development (18) in the Kirchhoff integral (15) we obtain

$$u_P = \frac{i\{\cos(nr) - \cos(nr_q)\}}{2\lambda RR_q} e^{-ik(R+R_q)} \int e^{-ik\phi(\xi, \eta)} dS. \quad (19)$$

### 3. Fraunhofer Diffraction by a Slit and by One-, Two- and Three-dimensional Gratings.

Following our classification of diffraction phenomena we imagine the source as well as the observer to be very far from the diffracting aperture, so that the waves which strike this aperture are essentially plane. This condition may be realized in practice by placing the source of light at the focus of a lens and by studying the diffraction pattern in the focal plane of another lens (see Chap. XXI).

As a first example, we study diffraction by a slit of width  $a$  which is of unlimited length in the  $\eta$ -direction. This means that the quantities occurring are independent of the  $y$ -co-ordinate, so that the problem is a two-dimensional one in the  $zx$ -plane. We also write  $i$  and  $\theta$  for the angles which an incident and a diffracted ray make with the normal  $Oz$  to the slit; hence  $\alpha_0 = \sin i$ , and  $\alpha = \sin \theta$ .

The diffraction integral reduces to

$$\begin{aligned} u_P &= \text{const.} \int_{-a/2}^{a/2} e^{-ik(\alpha_0 - \alpha)\xi} d\xi \\ &= \text{const.} \int_{-a/2}^{a/2} e^{-ik(\sin i - \sin \theta)\xi} d\xi. \quad \dots \quad (20) \end{aligned}$$

The value of the definite integral may be written down at once, but an instructive way to evaluate it is to use the vector diagram, and proceed graphically. We divide the slit into small strips of width  $d\xi$  and add the contributions of the separate strips. While  $\xi$  increases from  $-a/2$  to  $a/2$  in steps of amount  $d\xi$ , the phase difference increases in steps of  $k(\sin i - \sin \theta)d\xi$ . This means that the vectors representing two successive strips differ in direction by this amount. Since the contributions of all strips of equal width  $d\xi$  are of the same absolute magnitude, the series of index points form the vertices of a portion of a regular polygon. In the limit, where  $d\xi$  is infinitely small, the figure becomes a circular arc (cf. p. 52). The figure forms a closed loop when the phase difference between the first and the last strip is  $2\pi$ . A closed polygon means that the resultant disturbance is zero, i.e. there is complete darkness. In general, there is a minimum whenever

$$k(\sin i - \sin \theta)a = 2\pi n, \quad (n = 1, 2, 3, \dots)$$



or 
$$(\sin i - \sin \theta) a = n\lambda. \quad . \quad . \quad . \quad . \quad (21)$$

This is the condition for a minimum for diffraction by a single slit. Maxima occur between these minima at points where the resultant in the vector diagram is a diameter of the circle, i.e. where

$$k(\sin i - \sin \theta) a = (2n + 1)\pi, \quad (n = 1, 2, 3, \dots)$$

or where

$$(\sin i - \sin \theta) a = \frac{2n + 1}{2} \lambda. \quad . \quad . \quad . \quad . \quad (22)$$

This, then, is the condition for directions of maximum illumination in the diffraction pattern of a single slit.

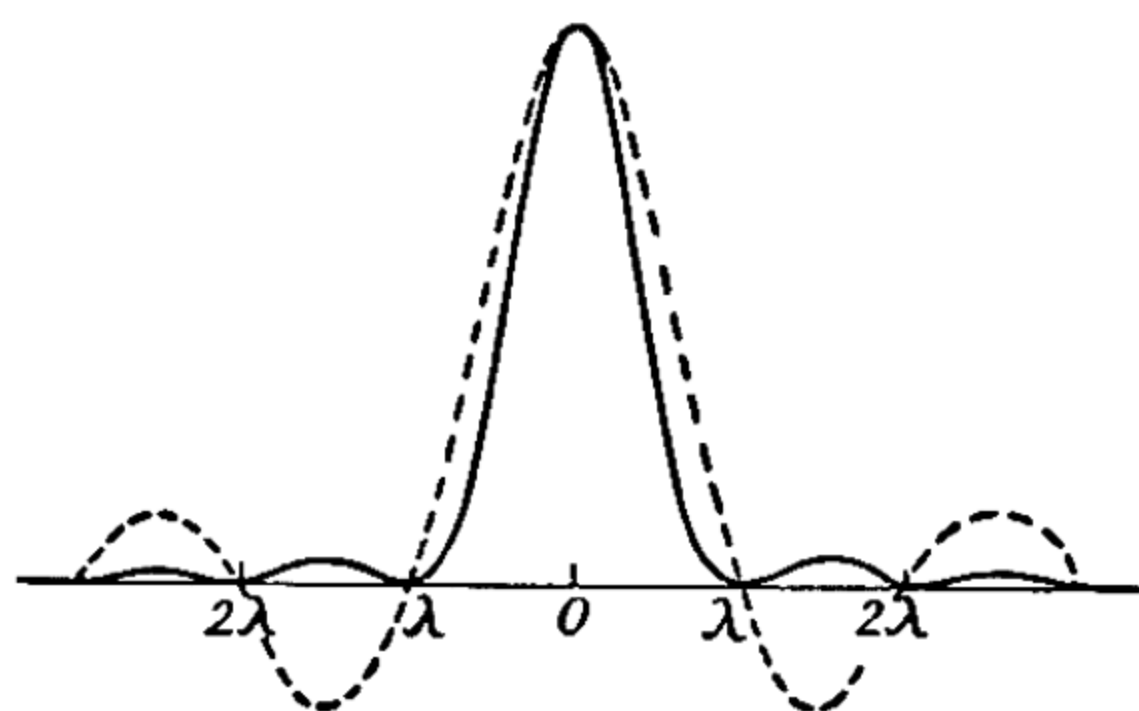


Fig. 3

For a given direction of incidence—which we may choose, for example, along  $Oz$  normal to the plane of the slit—the amplitude of the diffracted light will be a function of the difference in path for the two edges of the slit. The distribution of intensity may be found from (20), or from the vector diagram by the method of p. 53.

The result is shown in fig. 3, where the dotted curve gives the amplitude of the resultant disturbance, while the intensity is represented by the curve in full line.

If half the slit were covered there would be light in the directions

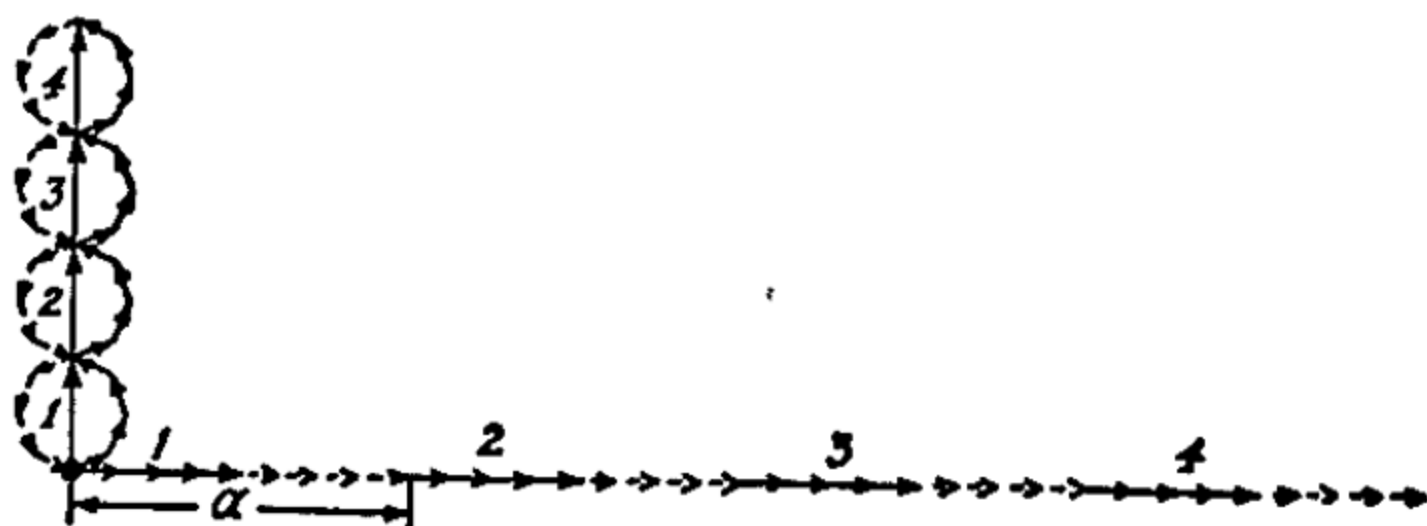


Fig. 4

of the previous minima when  $n$  is odd, since the second half of the index circle would now be missing. If we place a large number of such half-slits side by side we obtain a one-dimensional plane grating. Let the distance between corresponding points in two neighbouring openings be called  $a$ . The total effect is obtained by addition of the disturbances from the individual slits. In those directions for which

$$(\sin i - \sin \theta) a = m\lambda, \quad . \quad . \quad . \quad . \quad (23)$$

the individual disturbances plainly have the same phase, and hence there is a maximum of illumination (cf. fig. 4). This is then the condition for a maximum for the diffraction grating.

Since, for a given direction of the incident light, the angles giving maximum illumination depend upon the wave-length, a grating of this kind is eminently suitable for the spectral analysis of light. The number  $m$  is called the *order* of the spectrum. In what directions are the dark places to be found? From the vector diagram it is seen that the first dark spot occurs when the index points of the individual rulings (or slits) form a closed polygon. This is the case when the phase difference between the first and the last ruling is  $2\pi$ . Thus the rays coming from two adjacent lines of the grating must differ in phase by  $(2\pi m + 2\pi/N)$  to give the dark spot in the neighbourhood of the maximum of order  $m$ , where  $N$  is the total number of lines in the grating. Hence the direction of the minimum adjacent to the  $m$ th-order maximum is given by

$$(\sin i - \sin \theta)a = m\lambda + \frac{\lambda}{N}. \quad . \quad . \quad . \quad . \quad (24)$$

This is, then, the condition for points of minimum illumination in the diffraction pattern of a grating. We have here a means of computing the resolving power of a grating, i.e. the smallest relative difference in wave-length which can be separated. Following Rayleigh, two wave-lengths are considered separate if the maximum of the diffraction pattern of one coincides with the first minimum of the other. Let the wave-length  $\lambda + d\lambda$  have a maximum in the direction  $\theta$ , i.e.

$$(\sin i - \sin \theta)a = m(\lambda + d\lambda).$$

At the same time, suppose the first minimum of the wave-length  $\lambda$  occurs at the same point:

$$(\sin i - \sin \theta)a = m\lambda + \frac{\lambda}{N}.$$

From these two relationships we obtain

$$\frac{d\lambda}{\lambda} = \frac{1}{mN}.$$

The *resolving power*  $R$  is usually taken to be the reciprocal of this value, i.e.

$$R = \frac{\lambda}{d\lambda} = mN. \quad . \quad . \quad . \quad . \quad (25)$$

The directions of the maxima and minima obviously do not depend upon how the resultant of the disturbances from the *individual*

elements of the grating is formed. The only essential thing is that we have transparent or reflecting or scattering elements arranged at equal distances from one another. It is only the distribution of light among the various orders which is affected by the nature of the grating elements.

We pass now to the consideration of what we may call a two-dimensional grating. This consists of a large number of apertures whose distance apart is uniformly  $a$  in the  $\xi$ -direction and  $b$  in the  $\eta$ -direction. The conditions for maxima are obtained by the same construction as above, and are

$$\left. \begin{aligned} (a_0 - a)a &= m_1\lambda \\ (\beta_0 - \beta)b &= m_2\lambda \end{aligned} \right\} \cdot \cdot \cdot \cdot \cdot \cdot (26)$$

For a given direction of the incident light, the direction of the maximum corresponding to the order numbers  $m_1$  and  $m_2$  is uniquely determined, for the direction cosines  $a$ ,  $\beta$  each determine a cone about the corresponding axis. The generators common to these two cones give the directions of the maxima corresponding to the orders  $m_1$  and  $m_2$ . Addition of a third condition—as in the case of the space lattice (see below)—would define a third cone about the  $z$ -axis. This cone, however, would not necessarily contain the common generators of the two other cones.

While the two-dimensional grating has no practical application, the three-dimensional grating, or space lattice, is of the greatest importance—viz. in the analysis of electromagnetic waves of very short wave-length (X-rays). For such short waves it is not possible to produce appreciable diffraction angles with ordinary ruled gratings except by the use of special artifices. But nature has provided us with other gratings whose divisions are very much smaller, viz. crystals. The atoms in crystals are arranged at regular intervals, but form a three-dimensional lattice. In 1912 M. von Laue suggested the use of crystals as diffraction gratings and the experiment was performed successfully by Friedrich and Knipping.\*

We limit our considerations to rhombic lattices, in which the lattice points lie in three mutually perpendicular directions at distances  $a$ ,  $b$  and  $c$ . This, of course, includes the simpler gratings in which two or all three of the separations might be equal. The atoms act as sources of secondary waves, since they are set into vibration by the incoming light. Thus the secondary waves in this case have a certain reality, as contrasted with the Huygens wavelets in free space. Since the atoms are set into vibration by the incident light, they vibrate in phase, and the secondary waves which they produce (or scatter) are

\* M. von Laue, *Phys. Zeits.* 14, p. 421 (1913). W. Friedrich, P. Knipping and M. von Laue, *Le Radium*, 10, p. 47 (1913).



in phase, just as are ordinary Huygens wavelets. Hence the same treatment is applicable here. We now have three conditions for the directions of maximum illumination:

$$\left. \begin{aligned} (a_0 - a)a &= m_1\lambda \\ (\beta_0 - \beta)b &= m_2\lambda \\ (\gamma_0 - \gamma)c &= m_3\lambda \end{aligned} \right\}, \quad . . . . . (27)$$

or

$$\left. \begin{aligned} (a_0 - a) &= \frac{m_1\lambda}{a} \\ (\beta_0 - \beta) &= \frac{m_2\lambda}{b} \\ (\gamma_0 - \gamma) &= \frac{m_3\lambda}{c} \end{aligned} \right\}, \quad . . . . . (27')$$

or

$$\left. \begin{aligned} a &= a_0 - \frac{m_1\lambda}{a} \\ \beta &= \beta_0 - \frac{m_2\lambda}{b} \\ \gamma &= \gamma_0 - \frac{m_3\lambda}{c} \end{aligned} \right\}. \quad . . . . . (27'')$$

As we saw from the cone construction for the two-dimensional grating (p. 385), the addition of the third condition overdetermines the problem. This becomes evident mathematically by noticing that if we select  $a_0$ ,  $\beta_0$  and  $\gamma_0$  arbitrarily, the values of  $a$ ,  $\beta$ ,  $\gamma$  obtained from (27'') will not, in general, satisfy the condition  $a^2 + \beta^2 + \gamma^2 = 1$ . Hence, if we are to have a diffraction maximum at all, there must be a relationship between the direction and wave-length of the incident wave. If we square equations (27'') and add them, we obtain the required relationship:

$$\lambda = 2 \frac{\frac{m_1}{a}a_0 + \frac{m_2}{b}\beta_0 + \frac{m_3}{c}\gamma_0}{\frac{m_1^2}{a^2} + \frac{m_2^2}{b^2} + \frac{m_3^2}{c^2}}. \quad . . . . . (28)$$

A maximum is obtained only when this condition is fulfilled. We obtain a formula which lends itself more readily to direct interpretation if we square and add equations (27'). Denoting the angle between the directions of incidence and diffraction by  $\theta$  and remembering that

$$\cos \theta = aa_0 + \beta\beta_0 + \gamma\gamma_0,$$

we obtain

$$2(1 - \cos \theta) = \lambda^2 \left( \frac{m_1^2}{a^2} + \frac{m_2^2}{b^2} + \frac{m_3^2}{c^2} \right),$$

or

$$2 \sin \frac{\theta}{2} = \lambda \sqrt{\frac{m_1^2}{a^2} + \frac{m_2^2}{b^2} + \frac{m_3^2}{c^2}}. \quad \dots (29)$$

If the three order integers contain a common factor  $M$ , we place it outside the radical and have

$$2 \sin \frac{\theta}{2} = \lambda M \sqrt{\frac{H_1^2}{a^2} + \frac{H_2^2}{b^2} + \frac{H_3^2}{c^2}}, \quad \dots (30)$$

where  $H_1, H_2, H_3$  are relatively prime.

Planes may be passed through the crystal in various directions such that they contain a large number of lattice points (cf. fig. 5). A plane of this kind is known as a *reflecting plane*. A family of parallel reflecting planes is characterized by the set of relatively prime integers proportional to the reciprocals of the intercepts on the axes, each being divided by the corresponding grating constant. For example, (010) denotes the set of planes  $y = b, y = 2b, y = 3b, \&c.$  These indices, which are the  $H_1, H_2, H_3$ ,

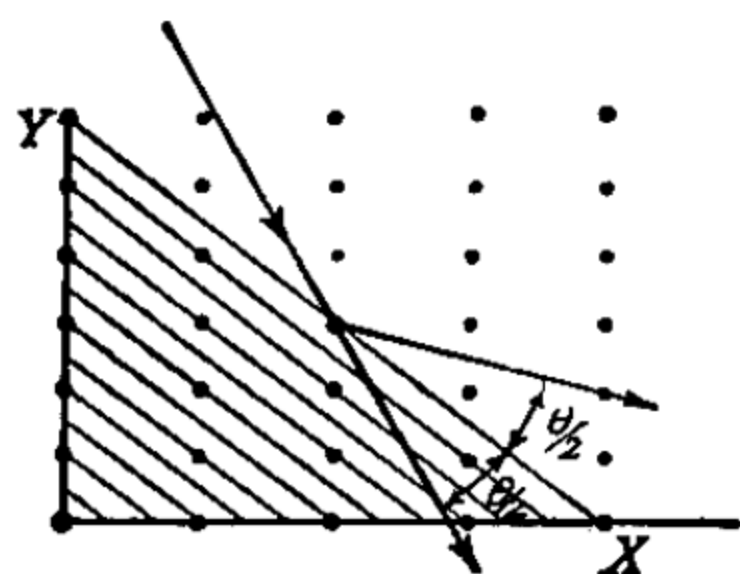


Fig. 5

represent the reciprocals of the actual intercepts of that plane which is nearest the origin. This may be seen as follows: If the intercepts of one of the planes are  $l_1, m_1, n_1$ , measured in centimetres, then

$$H_1 : H_2 : H_3 = \frac{a}{l_1} : \frac{b}{m_1} : \frac{c}{n_1},$$

or after introducing a factor of proportionality  $t$ ,

$$l_1 = \frac{ta}{H_1}, m_1 = \frac{tb}{H_2}, n_1 = \frac{tc}{H_3}.$$

The usual form of the equation of a plane whose axial intercepts are  $l_1, m_1, n_1$  is

$$\frac{x}{l_1} + \frac{y}{m_1} + \frac{z}{n_1} = \frac{xH_1}{ta} + \frac{yH_2}{tb} + \frac{zH_3}{tc} = 1.$$

If this is to be a reflecting plane, the equation must be satisfied by integral values of the co-ordinates, i.e. by  $x = pa, y = qb, z = rc$ . Inserting these values in the preceding equation, we have

$$pH_1 + qH_2 + rH_3 = t.$$

Since we have the sum of integers on the left,  $t$  must also be a whole number. Thus the smallest value of  $t$  is 1. According to the results of analytic geometry, the distance from the origin of a plane whose axial intercepts are  $a/H_1$ ,  $b/H_2$ ,  $c/H_3$  is

$$d = \frac{1}{\sqrt{\frac{H_1^2}{a^2} + \frac{H_2^2}{b^2} + \frac{H_3^2}{c^2}}}. \quad \dots \dots (31)$$

This is also the distance between two adjacent planes of the set. Thus we arrive at an entirely new interpretation of equation (30). There, after removing a common factor, the  $H_i$  represent the order numbers of the direction of diffraction. However, if we interpret the  $H_i$  as the indices of a family of planes, we can express the radicand in terms of the distance between the planes and thus obtain

$$2d \sin \frac{\theta}{2} = M\lambda. \quad \dots \dots (32)$$

This is the condition for a maximum in the diffraction pattern of a space lattice, and is known as Bragg's Law. The nature of these planes may be seen at once. It is easy to see that the plane half-way between the incident and diffracted rays (fig. 6) belongs to the set characterized by  $H_1, H_2, H_3$ .\* Hence, on account of the equality of the two angles, we may look upon the diffraction process as reflection of the rays by the reflecting planes. The difference in path of the parts of a wave reflected from the first plane and the succeeding plane is given by (see fig. 6)

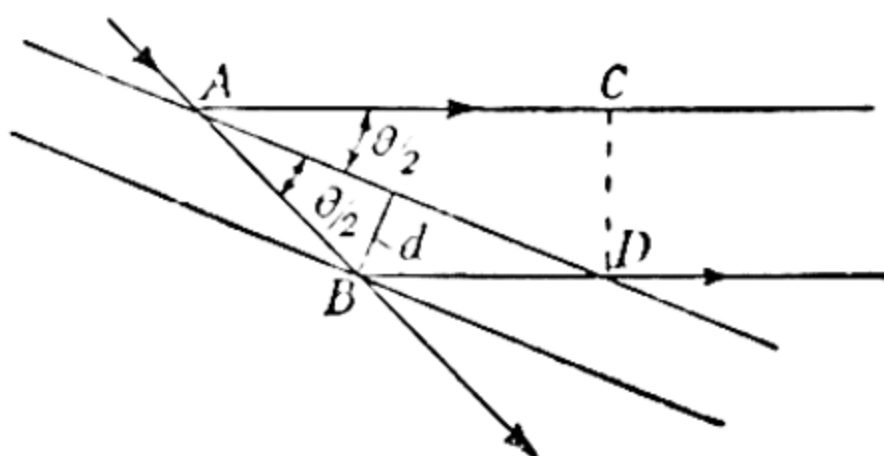


Fig. 6

$$AB + BD - AC = \frac{2d}{\sin \frac{\theta}{2}} - 2d \cot \frac{\theta}{2} \cos \frac{\theta}{2} = 2d \sin \frac{\theta}{2}. \quad (33)$$

Thus (32) signifies that for reflection from the individual planes the angle of incidence  $\pi/2 - \theta/2$ , the distance  $d$  between planes and the wave-length  $\lambda$  must fulfil the condition that the difference in phase between the partial rays must be  $2M\pi$ . Angles which satisfy equation

\* The equation of this plane is  $(s_0 - r)^2 = (s - r)^2$  or  $(s_0 - s)r = 0$ . If we insert for the components of  $(s_0 - s)$  the values given by equation (27') (p. 386) and remove the common factors, we obtain

$$\frac{H_1 x}{a} + \frac{H_2 y}{b} + \frac{H_3 z}{c} = 0.$$



(32) are called *glancing angles*. In order to obtain such a "reflection at a depth" with monochromatic radiation, it is necessary to rotate the crystal in such way that the entire range of angle between the reflecting planes and the incident wave normal is covered. This is the principle of the so-called Bragg Method.\*

Because the preceding derivation made use of a definite angle of incidence, it might be thought necessary to work with a parallel beam. For X-rays, such a beam could not be obtained by means of lenses, but solely by using two narrow slits. Meanwhile, Bragg and M. de Broglie discovered a focusing condition whose fulfilment guarantees that even a diverging bundle of rays from a single slit shall give a sharp line for each wave-length. The condition is (fig. 7) that the point of observation  $P$  and the slit  $S$  shall be equidistant from the axis of rotation  $O$ ,

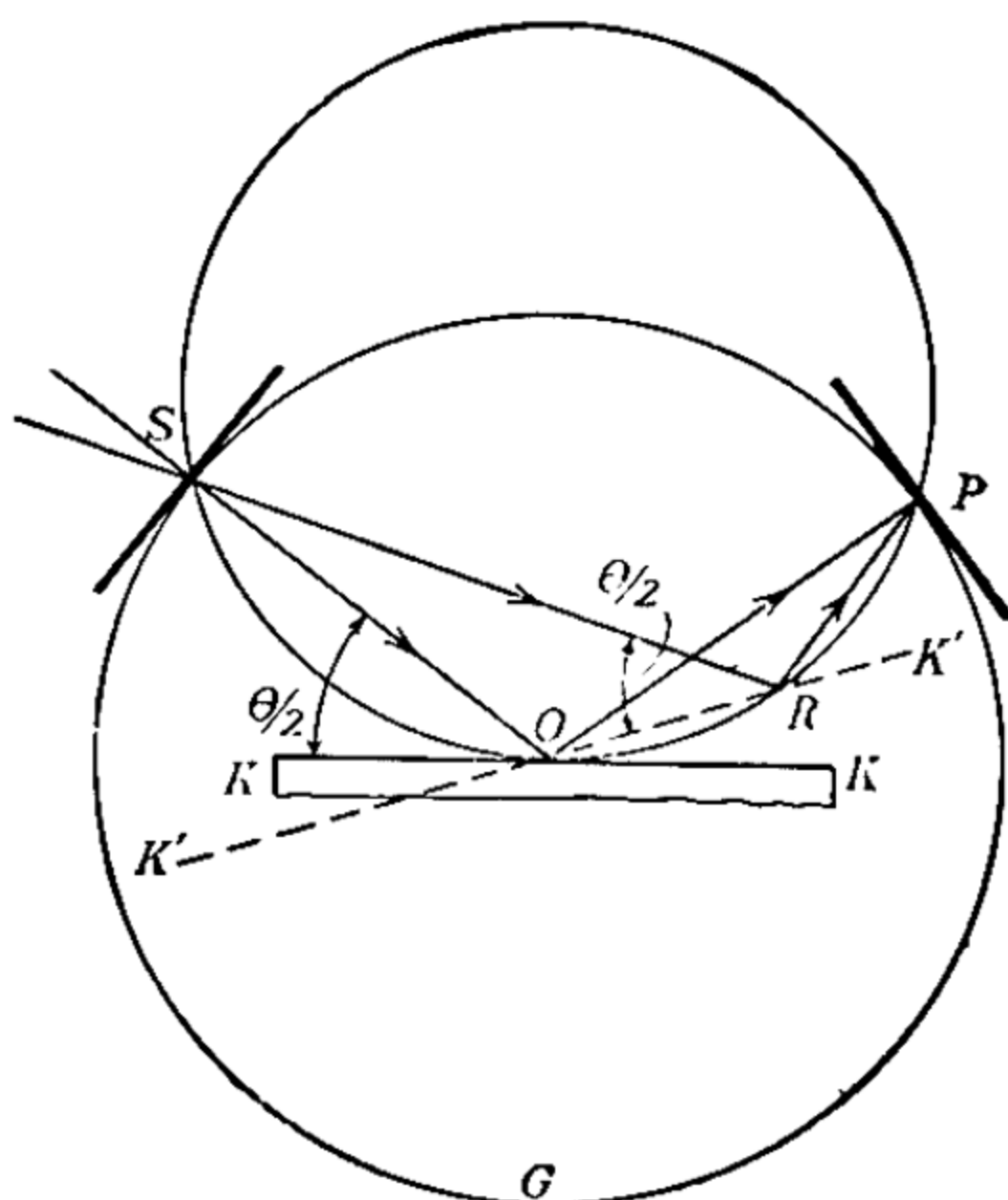


Fig. 7

i.e. the spectrum falls on a circle of radius  $SO$  drawn about  $O$ . In the symmetrical position  $KK$  of the crystal, a given wave-length reflected at  $O$  will arrive at  $P$ ; and for any other position  $K'K'$ , this same wave-length, and only this one, will go to the point  $P$ . If we draw a circle passing through  $S$ ,  $O$  and  $P$ , its intersection with  $K'K'$  will determine the point of reflection  $R$ . It is easy to see that  $SR$  and  $RP$  again make an angle of  $\theta/2$  with the surface of the crystal: angles  $SOP$  and  $SRP$  are equal, since they cut off the same arc. Also, angles  $SOK$  and  $SRO$  are equal since they are respectively the intercepting angle and the angle between tangent and chord corresponding to a given arc. Moreover, the latter angle is  $\theta/2$ . Then, also, angle  $POK$  is equal to angle  $PRK'$ .

It should be noticed that the advantage of this arrangement is not in increased intensity but in the circumstance that various parts of the crystal co-operate, and so chance local defects in structure become harmless.

Instead of turning a single crystal so that it assumes, successively, various

\* W. H. Bragg and W. L. Bragg, *X-Rays and Crystal Structure*, 5th Edition. London, G. Bell & Sons (1925).

positions with respect to the rays, we can assure incidence at all glancing angles by using a mass of small, randomly oriented crystals. This is the powder method devised by Debye and Scherrer and independently by A. W. Hull.\*

*Ex. 99.* Give the relative intensities of the diffraction patterns of various orders for a grating consisting of alternate transparent and opaque strips of equal width.

*Ex. 100. The Michelson Stellar Interferometer.*—Two parallel slits whose distance apart is  $b$  are cut from a screen. What is the appearance of the diffraction pattern if parallel rays are incident upon the screen (a) normally, and (b) at a small angle  $\epsilon$  with the normal? If both sets of rays are incident at the same time, how does the visibility of the pattern depend upon  $\epsilon$  and  $b$ ?

*Ex. 101.* A cubical crystal of KCl bounded by the planes (100), (010) and (001) is turned about an axis parallel to the line of intersection of the planes (100) and (010). The  $K_\alpha$ -radiation of copper ( $\lambda = 1541$  X.U.  $= 1541 \times 10^{-11}$  cm.) is incident normal to this axis. Which reflecting planes with indices  $\leq 2$  cause reflection, and what are the corresponding glancing angles  $\theta/2$ ? The constant of the cubic lattice is  $3.13$  A.U.  $= 3.13 \times 10^{-8}$  cm.

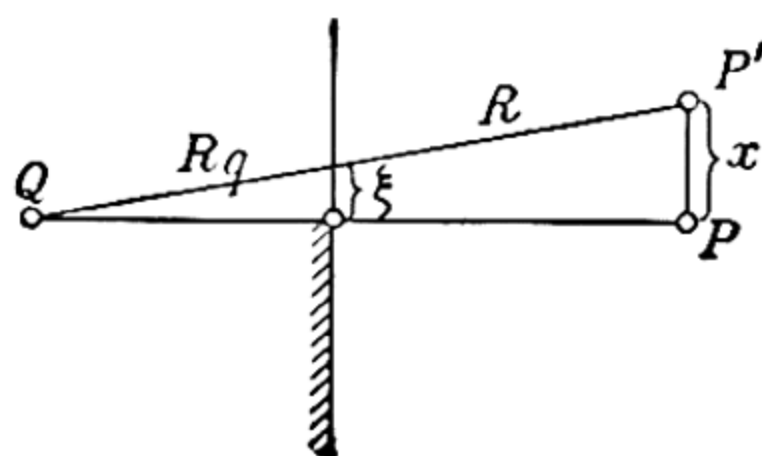


Fig. 8

#### 4. Fresnel Diffraction Phenomena at a Slit and at a Circular Aperture. Zone Plates.

Consider first a two-dimensional case. Imagine the source to be a line of light of indefinite extent in the direction normal to the plane of fig. 8, so that the  $y$ -co-ordinate does not enter. Further, let the perpendicular from the point of observation to the line source be normal to the plane of the screen, and let it pass through one edge of the slit. Since  $\sin i = \sin \theta = 0$ , i.e.  $\alpha_0 = \alpha = 0$ , in this case, the development of  $\phi(\xi, \eta)$  begins with

$$\frac{\xi^2}{2} \left( \frac{1}{R} + \frac{1}{R_q} \right)$$

(cf. p. 381). Replacing  $k$  by its value  $2\pi/\lambda$  (equation (4) (p. 377), and supposing the slit to extend from  $\xi = 0$  to  $\xi = a$ , we obtain for the disturbance at  $P$  the expression

$$u_P = \frac{i}{\lambda R R_q} e^{-i2\pi/\lambda \cdot (R+R_q)} \int_0^a e^{-i\pi/\lambda \cdot \xi^2 (1/R+1/R_q)} d\xi. \quad (34)$$

\* See G. I. Clark, *Applied X-Rays*, 3rd Edition, p. 267 et seq., New York, McGraw-Hill (1940).

If we evaluate the integral as before, by summing the "vectors" of individual elements of the slit, we see that in contrast with the case treated in § 3 (p. 382), the vectors of the several elements  $d\xi$  no longer make equal angles with one another; on account of the quadratic

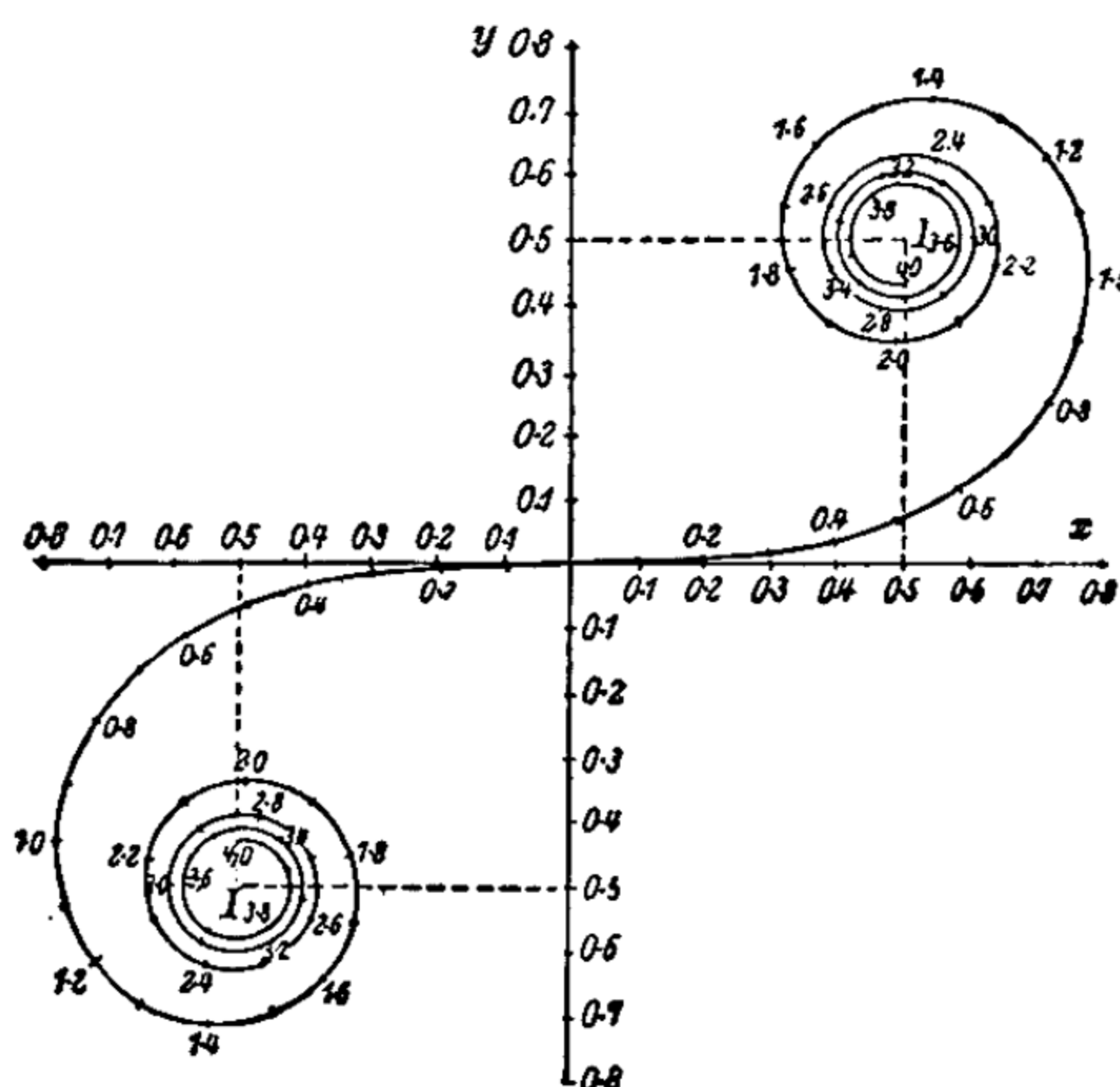


Fig. 9

term the angle between two consecutive vectors increases as  $\xi d\xi$ , i.e. the curvature of the vector curve increases with the length of arc. A curve of this nature has a spiral form. In order to obtain an integral independent of the special values of  $\lambda$ ,  $R$ ,  $R_q$  we make the substitution

$$s = \xi \sqrt{\frac{2}{\lambda} \left( \frac{1}{R} + \frac{1}{R_q} \right)}; \quad A = a \sqrt{\frac{2}{\lambda} \left( \frac{1}{R} + \frac{1}{R_q} \right)} \quad (35)$$

and obtain

$$\int_0^a e^{-i\pi/\lambda \cdot \xi^2 (1/R + 1/R_q)} d\xi = \frac{1}{\sqrt{\frac{2}{\lambda} \left( \frac{1}{R} + \frac{1}{R_q} \right)}} \int_0^A e^{-i\pi/2 \cdot s^2} ds. \quad (36)$$

The curve

$$z = \int_0^A e^{i\pi/2 \cdot s^2} ds$$

whose parametric equations in rectangular co-ordinates are

$$x = \int_0^A \cos\left(\frac{\pi}{2} s^2\right) ds, \quad y = \int_0^A \sin\left(\frac{\pi}{2} s^2\right) ds \quad (37)$$

is called the Cornu Spiral (fig. 9). The numbers marked on the curve



are the values of  $A$ . Each number also gives the length of the corresponding arc of the curve measured from  $O$ ; this is seen at once, because we have  $dx/dA = \cos \pi A^2/2$ , and  $dy/dA = \sin \pi A^2/2$ , so that  $(dx/dA)^2 + (dy/dA)^2 = 1$  or  $ds/dA = 1$ , where  $s$  is the length of the arc.

Once we have drawn this spiral we can employ it to answer a number of questions.

First we notice that the radius vector from  $O$  to the point  $A$  of the spiral, whose relation to the slit width  $a$  is given by (35), gives the disturbance at  $P$  as a function of the slit width (with the sign of the imaginary part changed). As the figure shows, this radius vector passes through infinitely many turning values, whose difference becomes smaller and smaller, the value approaching that of the asymptotic point  $I$  when the slit width becomes infinite, i.e. the slit becomes a single diffracting edge. If, now, the half of the slit symmetric to the above is opened also, an equally great light disturbance must be added. For convenience we draw the spiral corresponding to the other half so that it is centrally symmetric about  $O$  with respect to the original spiral. It must be remembered, however, that for the second spiral the radius vectors are to be drawn *from* points on the curve *to* the origin. We are now in a position to give the value of the disturbance at  $P$  when the slit has any position whatsoever on the  $\xi$ -axis—for example, when its edges are at  $\xi = a$  and  $\xi = b$ . We need only subtract the vector corresponding to  $\xi = b$  from that of  $\xi = a$ , i.e. draw the line connecting the points on the curve which correspond to  $a$  and  $b$ .

Further, we can calculate with good approximation the distribution of light on a horizontal line through  $P$  parallel to the screen. If we limit our inquiry to points not too far from  $P$  we may neglect the small inclination of the lines  $QP'$  and we may expect the same disturbance at  $P'$  as would exist at  $P$  if the screen ended at  $\xi = -xR_q/(R + R_q)$  instead of at  $\xi = 0$  (cf. fig. 8). With  $x = PP'$  increasing in negative values we traverse the points of the upper spiral, and the distances from the asymptotic point to these points steadily decrease. Hence the illumination in the region of shadow declines uniformly. Displacing  $P$  by a positive  $x$  corresponds to withdrawing the edge of the screen in the negative direction. In this case the lower spiral also contributes radius vectors, giving a series of maxima and minima in the region of illumination. The limit for large distance from the geometric shadow corresponds to the line joining the two asymptotic points, i.e. double the amplitude (or four times the intensity) at the point  $P$ . This distribution of light corresponds well with observation.

Finally, we wish to investigate the distribution of light along a normal to the screen when the source of light is a *point*.

If we again have  $a_0 = a = 0$ , the disturbance at  $P$  is given by

$$u_P = \frac{i}{\lambda R R_q} e^{-ik(R+R_q)} \iint e^{-ik/2 \cdot (\xi^2 + \eta^2)(1/R + 1/R_q)} d\xi d\eta. \quad (38)$$

For a circular opening (radius  $a$ ) the integral can be evaluated, and we obtain

$$u_P = \frac{1}{R + R_q} e^{-ik(R+R_q)} \{ 1 - e^{-i\pi/\lambda \cdot a^2(1/R + 1/R_q)} \}. \quad (39)$$

The intensity is given by the square of the modulus of the complex number, and this is given by multiplication by the conjugate complex number. Also

$$(1 - e^{-ix})(1 - e^{+ix}) = 2(1 - \cos x) = 4 \sin^2 \frac{x}{2}.$$

The result is

$$I_P = \frac{4}{(R + R_q)^2} \sin^2 \frac{\pi}{2\lambda} a^2 \left( \frac{1}{R} + \frac{1}{R_q} \right). \quad (40)$$

Thus we see that for increasing size of aperture the illumination varies periodically. It must be borne in mind, however, that the formula is not to be applied to arbitrarily large apertures, since the inclination of the directions of  $r_0$  and  $r$  is of importance in such cases. A maximum of intensity occurs when

$$\frac{1}{R} + \frac{1}{R_q} = (2m + 1) \frac{\lambda}{a^2}. \quad (41)$$

If we compare this formula with the elementary lens formula, we see that the circular aperture corresponds to a convex lens of focal length  $f = a^2/(2m + 1)\lambda$ . However, we find no appreciable increase in intensity at the maxima here; this is first realized by diaphragming out those zones whose phase is such as to cause a decrease of the resultant amplitude. This corresponds to rendering alternate strips of a slit opaque, i.e. to the passage to a line grating. The division into zones in the present case must be such that the edges of adjacent zones, of which one half is covered, furnish a phase difference of  $2m\pi$ . Thus we must have

$$\frac{\pi(\rho_{n+1})^2}{\lambda} \left( \frac{1}{R} + \frac{1}{R_q} \right) = \frac{\pi\rho_n^2}{\lambda} \left( \frac{1}{R} + \frac{1}{R_q} \right) + 2m\pi,$$

or

$$\rho_{n+1}^2 - \rho_n^2 = \frac{2m\lambda}{\frac{1}{R} + \frac{1}{R_q}}. \quad (42)$$

It is seen that this also satisfies the condition that each zone has the same area, since this quantity is given by  $\pi(\rho_{n+1}^2 - \rho_n^2)$ . The innermost circle, half of whose area is to be covered, has a radius

$$\rho_1 = \sqrt{\frac{2m\lambda}{\frac{1}{R} + \frac{1}{R_q}}}$$

With a plate so divided into alternate transparent and opaque rings we can actually obtain a considerable intensification of light at points which satisfy the equation

$$\frac{1}{R} + \frac{1}{R_q} = \frac{2m\lambda}{\rho_1^2}.$$

For this reason the so-called *zone plate* is sometimes referred to as a *zone lens*.



## CHAPTER XXI

### THE ELEMENTS OF GEOMETRICAL OPTICS AND OF INTERFERENCE OPTICS

#### 1. The Fundamentals of Geometrical Optics. Laws of Fermat and of Malus.

Geometrical optics has developed as a special field serving the needs of practical optics. Its purpose is to compute the details of the passage of light through optical instruments and to determine the principles of the construction of such instruments. The fundamental hypotheses of this branch of optics represent a great simplification of the actual facts. They are essentially the following:

1. Rectilinear propagation of light rays in a homogeneous medium.
2. Independence of the several rays.
3. Reversibility of the rays.
4. The Law of Reflection.
5. The Law of Refraction (Snell's Law).

Assumptions 1, 2 and 3 show that geometrical optics not only takes no account of diffraction, but that it does not even contain any reference to the wave nature of light. Indeed, in the realm of diffraction phenomena the concept of a light ray loses all meaning; while it is possible to assign a ray in the direction of the normal to a plane wave incident on an aperture, the normals corresponding to the parts of the diffracted wave have very diverse directions. In the calculation of optical instruments with the aid of geometrical optics, the phenomena due to diffraction at the various apertures or diaphragms are not taken into consideration. When needed—e.g. for the determination of resolving power—these diffraction effects must be treated separately by the methods of the Theory of Diffraction.

Two important general theorems follow from Hypotheses 1, 4 and 5. The first of these was discovered by Fermat and is stated thus: The propagation of light always takes place in such way that the actual *optical path* (length of geometric path multiplied by index of refraction of the medium) is an extremum (i.e. has a stationary value) compared with all other paths which do not follow the laws of optics, i.e.

$$\Sigma n, s_1 = \text{Extremum.} \quad . \quad . \quad . \quad . \quad . \quad (1)$$

For media in which the index varies continuously, this becomes

$$\int_{P_1}^{P_2} n ds = \text{Extremum.} \quad . \quad . \quad . \quad . \quad (1')$$

To prove this we shall show that the first variation of the optical path vanishes. For propagation in a homogeneous medium the law is self-evident, since the straight line is the shortest distance between two

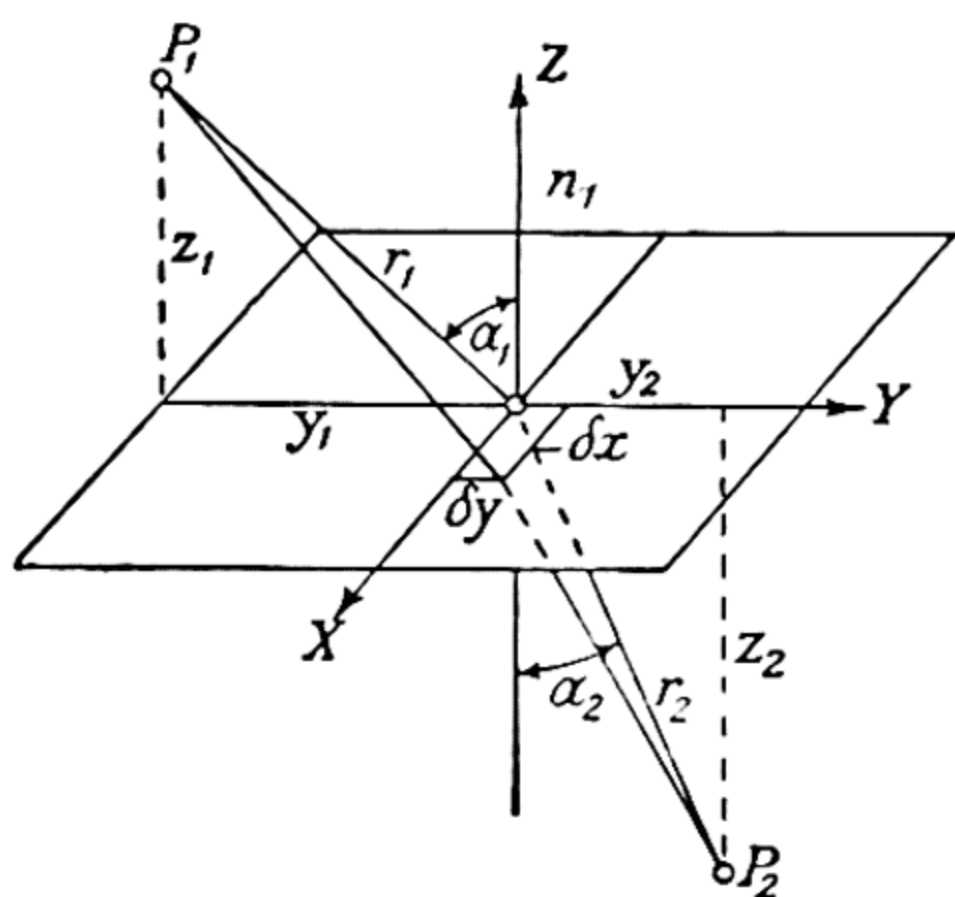


Fig. 1

points. Again, for the case of reflection, the proof may be reduced to this first case, since if the laws of geometrical optics are valid, the light reaching  $P_2$  may be considered to come from the image  $P_1'$ ; the path  $P_1'P_2$  is equal to the actual path and, being a straight line, is again an extremum. We give a proof also for the case of refraction. Let the refracting surface be the  $xy$ -plane of a co-ordinate system and let the plane of incidence be the  $yz$ -plane. Besides the actual path we consider a varied path (fig. 1)

whose intersection with the  $xy$ -plane has the co-ordinates  $\delta x$  and  $\delta y$ , while the "optical" ray passes through  $O$ . Then, if  $L$  is the optical path, we have

$$\begin{aligned} \delta L &= n_1(r_1 + \delta r_1) + n_2(r_2 + \delta r_2) - (n_1 r_1 + n_2 r_2) \\ &= n_1 \sqrt{z_1^2 + (y_1 - \delta y)^2 + \delta x^2} + n_2 \sqrt{z_2^2 + (y_2 - \delta y)^2 + \delta x^2} \\ &\quad - n_1 \sqrt{z_1^2 + y_1^2} - n_2 \sqrt{z_2^2 + y_2^2}. \quad . \quad . \quad . \quad . \quad (2) \end{aligned}$$

Neglecting terms of second order, this becomes simply

$$\delta L = - \left( \frac{n_1 y_1}{\sqrt{z_1^2 + y_1^2}} + \frac{n_2 y_2}{\sqrt{z_2^2 + y_2^2}} \right) \delta y.$$

But

$$\frac{n_1 y_1}{\sqrt{z_1^2 + y_1^2}} + \frac{n_2 y_2}{\sqrt{z_2^2 + y_2^2}} = -n_1 \sin \alpha_1 + n_2 \sin \alpha_2.$$

However, by Snell's Law, the right member is zero, which proves the theorem for refraction. Since any optical path in geometrical optics may be considered to be a combination of reflected and refracted rays, the theorem is thus demonstrated in the general case.

If an image of a point  $P$  is formed at another point  $P'$  many rays unite at the image, all of them having travelled according to the laws of geometrical optics. Each ray must represent a stationary optical

path, as compared with a neighbouring path which does not obey these laws. This is only possible if the optical path is constant for all optical rays. From the standpoint of wave optics, the constancy of optical path is a basic necessity, since the waves which converge in the image must all arrive in the same phase.

In order to understand the second important general theorem, the Law of Malus, it must be recalled that we distinguish between two kinds of ray systems in geometry—those which represent the normals to a surface (orthotomic systems of rays) and those having no orthogonal surfaces (non-orthotomic systems). The rays emanating from a point are certainly orthotomic, for the spheres with centre at this point are the required orthogonal surfaces. The Law of Malus states that the property of orthotomy is not destroyed by reflection or refraction, although the originally spherical orthogonal surfaces may be altered in shape. This law is self-evident from the standpoint of wave optics, for there must always be wave fronts corresponding to the rays. We therefore refrain from giving a proof from the point of view of geometrical optics. One further consequence of the Law of Malus is important. If we consider the parts of two rays included between two surfaces, their optical paths must be equal in length. This must also be true for rays that intersect later on, in which case the second wave surface passes through the point of intersection, where it has a singular point.

## 2. The Properties of Collinear Projection.

The ideal case of optical image formation is that in which all rays coming from a point unite again at another point, while the images of straight lines are straight lines, and the images of planes are again planes. *A priori*, one cannot say whether or not this type of image formation is possible with optical means. We shall assume that it is possible, and without going deeper into the question we investigate the properties of this type of correspondence between the points of the *object space* and those of the *image space*. Such a correspondence between the co-ordinates  $\xi, \eta, \zeta$ , of  $P$  and  $\xi', \eta', \zeta'$  of  $P'$  is represented mathematically by linear fractional transformations

$$\left. \begin{aligned} \xi' &= \frac{a_1\xi + b_1\eta + c_1\zeta + d_1}{a_0\xi + b_0\eta + c_0\zeta + d_0} \\ \eta' &= \frac{a_2\xi + b_2\eta + c_2\zeta + d_2}{a_0\xi + b_0\eta + c_0\zeta + d_0} \\ \zeta' &= \frac{a_3\xi + b_3\eta + c_3\zeta + d_3}{a_0\xi + b_0\eta + c_0\zeta + d_0} \end{aligned} \right\} \cdot \cdot \cdot (3)$$



As may be verified immediately, the equation of a plane

$$l\xi' + m\eta' + n\zeta' - p = 0$$

in the image space corresponds to a plane in the object space also. The intersection of two planes—i.e. a line—in the object space corresponds to a line in the image space, and the intersection of three planes—i.e. a point—again corresponds to a point. We shall restrict our discussion to axially symmetric (centred) systems.

If the  $\xi$ -axis is the axis of symmetry, we need investigate the image formation only in a meridional section, using the co-ordinates  $\xi$  and  $\eta$ . The symmetry of the image formation requires only that the  $\xi'$  co-ordinate be unchanged when  $\eta$  is replaced by  $-\eta$  but that the sign of  $\eta'$  change when that of  $\eta$  is reversed. Hence for a centred system  $b_0$ ,  $b_1$ ,  $a_2$  and  $d_2$  vanish, so that we now have the simpler expressions

$$\xi' = \frac{a_1\xi + d_1}{a_0\xi + d_0}, \quad \eta' = \frac{b_2\eta}{a_0\xi + d_0}. \quad \dots \quad (3')$$

Solved for  $\xi$  and  $\eta$ , these equations are

$$\xi = \frac{d_0\xi' - d_1}{-a_0\xi' + a_1}, \quad \eta = \frac{(a_1d_0 - a_0d_1)\eta'}{b_2(-a_0\xi' + a_1)}. \quad \dots \quad (3'')$$

The infinitely distant plane of the image space corresponds to the plane  $a_0\xi + d_0 = 0$ ; that of the object space corresponds to the plane  $a_0\xi' - a_1 = 0$ . These are called the *focal planes* of the respective regions. Their two points of intersection with the axis are called the *principal foci*  $F$  and  $F'$  of the system. The transformation formulæ become still simpler if we move the zero points of the  $\xi$ - and  $\xi'$ -axes to the respective foci. We shall also take the positive direction of the axes in object and image space to be that in which the light is propagated. Denote the co-ordinates of this system by  $x$  and  $y$  and  $x'$  and  $y'$  respectively. We then have the transformation

$$\left. \begin{aligned} x &= \xi + \frac{d_0}{a_0}, & y &= \eta \\ x' &= \xi' - \frac{a_1}{a_0}, & y' &= \eta' \end{aligned} \right\} \dots \quad (4)$$

Introducing the notation

$$\frac{b_2}{a_0} = -f, \quad \frac{a_0d_1 - a_1d_0}{a_0^2} = ff', \quad \dots \quad (5)$$

we then obtain

$$\left. \begin{aligned} &xx' = ff' \\ &\frac{y'}{y} = -\frac{f}{x} = -\frac{x'}{f'} \end{aligned} \right\} \dots \dots \dots (6)$$

The lengths  $f$  and  $f'$  introduced here are called the *focal lengths* of object space and image space respectively. We shall see at once that they are identical with the focal lengths introduced in the elementary lens theory.

The ratio  $y'/y$  is termed the *transverse magnification*. For the points of the plane  $x = -f$ , which corresponds to the plane  $x' = -f'$ , this ratio is unity. These two planes are called the *principal planes* or *unit planes*. Their points of intersection with the axis are the *principal points*  $H$  and  $H'$ . Any optical system may be schematically represented by

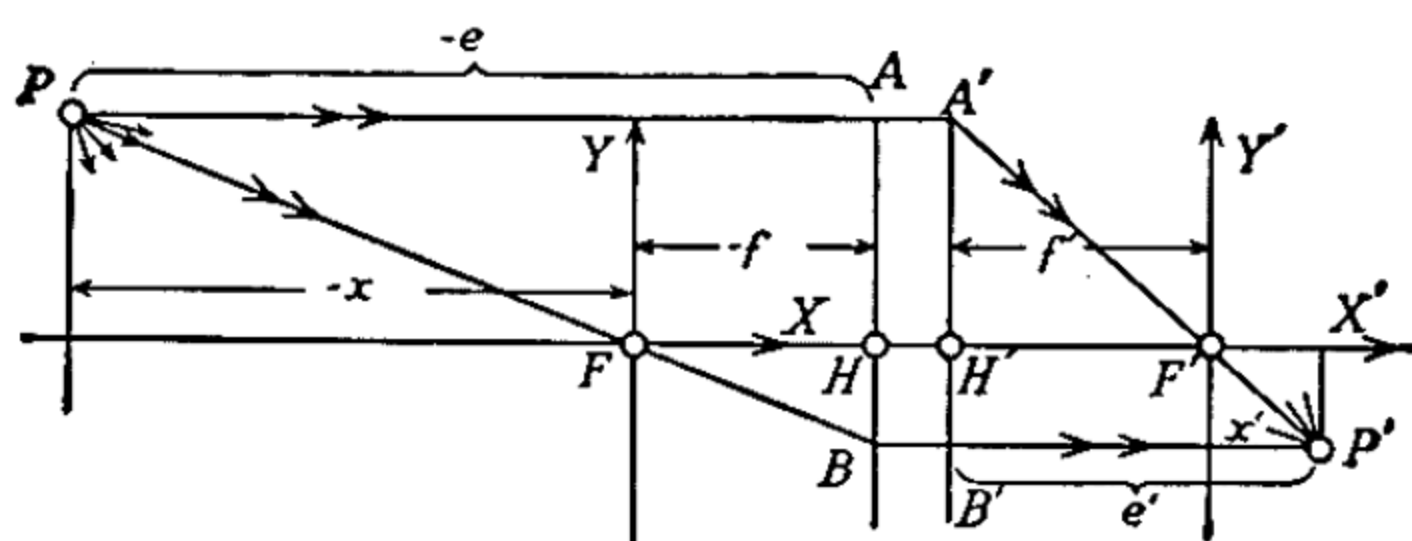


Fig. 2

an axis, the foci and the principal points. The distance  $e$  from a point of the object to the principal plane in the object space is called the *object distance*; similarly,  $e'$ , the *image distance*, is that of the corresponding image point from its principal plane. Then (see fig. 2),

$$-e = f - x. \quad e' = f' + x'$$

and the first equation of (6) gives

$$(e - f)(e' - f') = ff' \quad \text{or} \quad \frac{f}{e} + \frac{f'}{e'} = 1. \quad \dots \dots (7)$$

For a symmetric system,  $f = -f'$  and we obtain the elementary lens formula \*

$$-\frac{1}{e} + \frac{1}{e'} = \frac{1}{f}. \quad \dots \dots \dots (7')$$

In elementary optics we measure the image and object distances for a thin lens from the middle of the lens. For a more complicated optical system, e.g. a photographic objective, this is no longer allow-

\* This formula is usually written for the absolute values, i.e. without the minus sign attached to  $1/e$ . The same is true for equation (8), following.





Since this quantity depends upon  $x$ , we see that collinear projection does not preserve similarity.

If  $f$  and  $f'$  have different signs, as assumed, then a movement of the object will cause the image to move in the same direction. This is the case for reflection.

The *angular magnification* or *convergence ratio* is defined as the ratio of the tangents of the angles between the axis and a ray in image space and in object space, the ray being such as to cut the axis in each

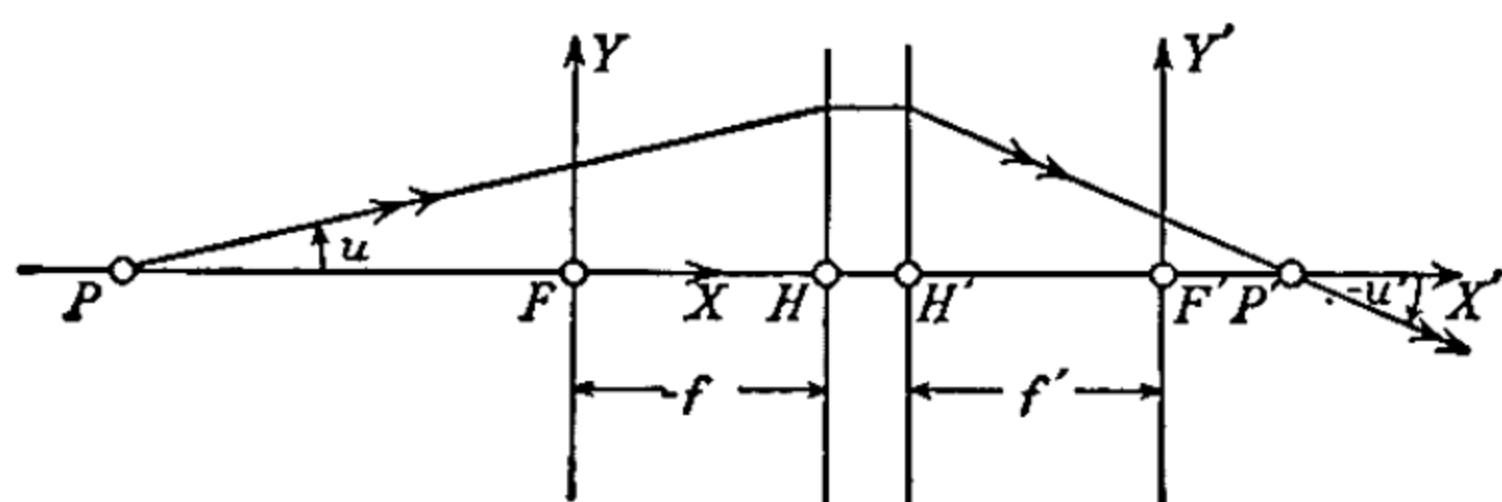


Fig. 4

region. This ratio is constant for a given point on the axis, for it follows from fig. 4 that

$$\frac{\tan u'}{\tan u} = \frac{f + x}{f' + x'} = \frac{f + x}{f' + \frac{ff'}{x}} = \frac{(f + x)x}{f'(x + f)} = \frac{x}{f'}. \quad (10)$$

The product of the transverse magnification and angular magnification for a given plane is constant for the entire system, for we have from equation (10) and equation (6) (p. 399):

$$\frac{y'}{y} \frac{\tan u'}{\tan u} = -\frac{f}{f'}. \quad (11)$$

### Combination of two imaging processes

The result of two successive collinear transformations is again collinear, and so there must be a single projection equivalent to the two. We set ourselves the task of determining the four cardinal points for the result, i.e. the focal points and the principal points or their corresponding planes. Fig. 5 shows these planes for each of the two processes, denoted respectively by subscripts 1 and 2. The same letters will be used to indicate the intersection of each plane with the axis. A ray parallel to the axis and distant  $y$  from it is drawn in the object space. After meeting  $H_1$  and  $H_1'$  it goes through the focus  $F_1'$ . Its inclination to the axis in the first image space is

$$\tan u_1' = -\frac{y}{f_1'}. \quad (12)$$

The final focal point will be the image of the object  $F_1'$ , as formed by the second system. Denoting the separation  $F_1'F_2$  by  $D$ , which is reckoned positive when  $F_2$  is to the right of  $F_1'$ , we have by equation (6), p. 399:

$$x_{2(F_1')} = -D, \quad x'_{2(F_1')} = \overline{F_2'F'} = -\frac{f_2 f_2'}{D}. \quad \dots (13)$$

The principal plane  $H'$  may be located graphically as follows: If  $H'$  is the principal plane of the combined system, a horizontal incident ray distant  $y$  from the axis will finally pass through  $F'$ . Then  $H'$  must contain the intersection of the horizontal line and the ray pass-

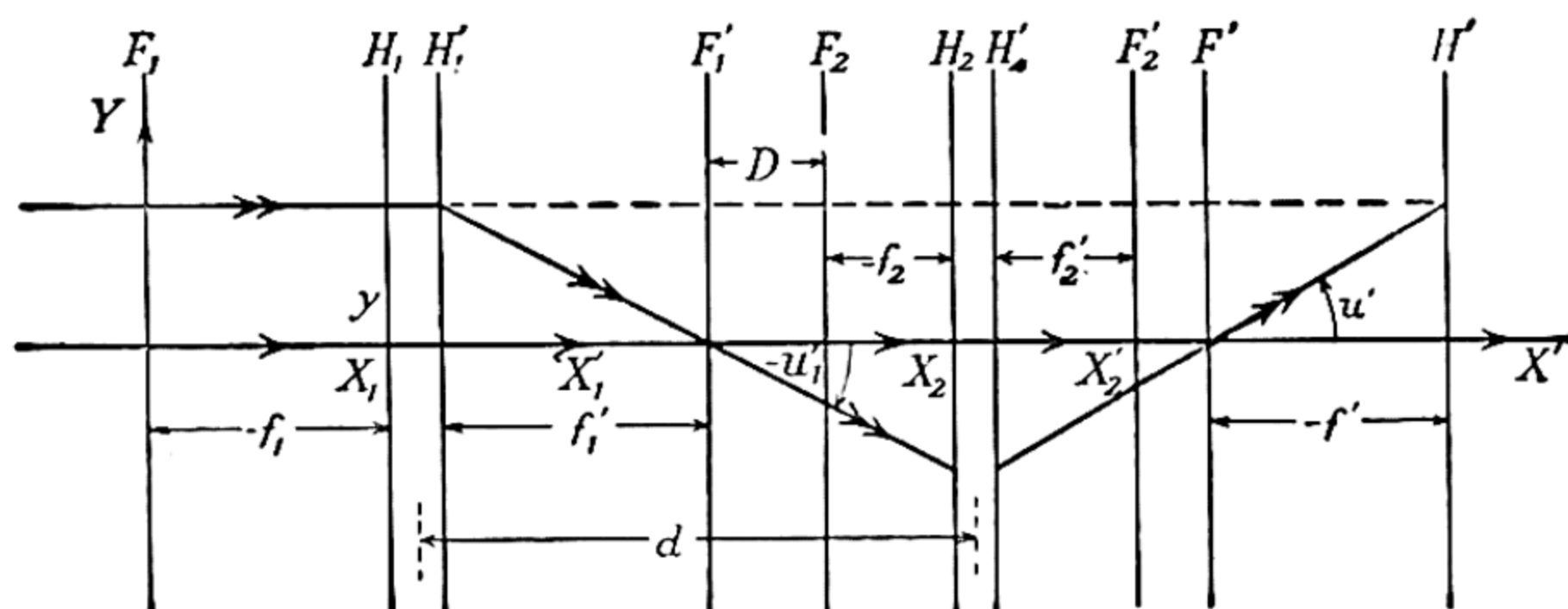


Fig. 5

ing through  $F'$  that corresponds to the ray  $F_1'H_1'$ . Algebraically, the focal length  $f'$ —the distance between  $F'$  and  $H'$ —is obtained by applying equation (10) to the angular magnification produced in the second system, remembering that  $u_2 = u_1'$ . Then

$$\tan u_2' = -\tan u_1' \cdot \frac{D}{f_2'} = \frac{y}{f_1'} \cdot \frac{D}{f_2'}.$$

Also, according to fig. 5, since  $F'$  is to the left of  $H'$ ,

$$\tan u' = \tan u_2' = -\frac{y}{f'},$$

whence

$$\frac{1}{f'} = -\frac{D}{f_1' f_2'} \quad \dots \dots \dots (14)$$

Similarly, one finds

$$F_1 F = \frac{f_1 f_1'}{D} \quad \text{and} \quad \frac{1}{f} = \frac{D}{f_1 f_2} \quad \dots \dots \dots (14')$$

If, as is usual, the same medium (air) is present on both sides of the lenses, then the two focal lengths will be equal in magnitude and opposite in sign. This will be shown in the next section. If the system

consists of two thin lenses, for each of which the principal planes may be assumed approximately coincident, the distance  $d$  between lenses may be written  $d = D + f_1' - f_2$ , or  $d = D + f_1' + f_2'$  (see figure). Putting this value of  $D$  in (14) and in (14') we get, with the lenses in air,

$$\frac{1}{f'} = \frac{1}{f_1'} + \frac{1}{f_2'} - \frac{d}{f_1'f_2'} \quad \text{and} \quad \frac{1}{f} = \frac{1}{f_1} + \frac{1}{f_2} + \frac{d}{f_1f_2}. \quad (15)$$

If, in addition, the lenses are close together, the result is simply

$$\frac{1}{f} = \frac{1}{f_1} + \frac{1}{f_2} \cdot \cdot \cdot \cdot \cdot \cdot \cdot \quad (16)$$

Hence for thin lenses in contact, the reciprocals of the separate focal lengths simply add up to give the reciprocal of the combined focal length. For this reason the reciprocal of the focal length of a lens is called the *power*  $\phi$  and is measured in *diopeters*, one diopter being the power of a lens whose focal length is one meter. It must be remembered that the additivity of the powers—which holds, incidentally, for any number of lenses—is valid only when the same medium is present on both sides of the system.

*Ex. 102.* The solar disc has an angular diameter, as seen from the earth, of  $1/2^\circ$ . How large is the image formed by a lens of focal length 2 m.?

*Ex. 103.* The plane of an object is inclined to the axis of a lens system. Locate the image plane.

### 3. The Practical Problem of Image Formation. Abbe's Sine Law. General Path of an Elementary Pencil from a Point Source.

#### (a) *Imaging of points near the axis by means of narrow pencils*

Suppose that the angle between the rays diverging from a point on the axis is so small that the sine and tangent may be taken equal to the angle itself. Assume further that only points near the axis are considered, so that their distances from the axis may be taken equal to the lengths of circular arcs whose centres are on the axis. Under these conditions, calculation of the refraction at a centred spherical surface leads to the collinear formulæ. Since the net effect of successive collinear transformations is again collinear and since, to the above approximation, any surface of rotation may be replaced by its sphere of curvature, it may be said in general that within the limits of accuracy described, *any system consisting of coaxial refracting surfaces of rotation is collinear*. This statement is true even if some of the surfaces are reflectors, since this case is merely one where  $n = -1$ .

The truth of the above statement will be shown for a refracting spherical surface separating a medium of index  $n$ , on the left, from one



of index  $n'$  on the right. The convention regarding algebraic signs will be that all axial segments to the left are negative, those to the right are positive; the curvature of a surface is positive in the convex side is facing the light. In fig. 6, let  $P$  be an axial point whose distance from the vertex of surface  $R$  is  $-s$ ; and let a ray leaving  $P$  at an angle  $\epsilon$  with the axis arrive, after refraction, at another axial point  $P'$  whose distance from  $R$  is  $s'$ . We must show that  $s'$  is independent of the angle  $\epsilon$ , i.e. that all rays coming from  $P$  again unite at  $P'$ . Draw

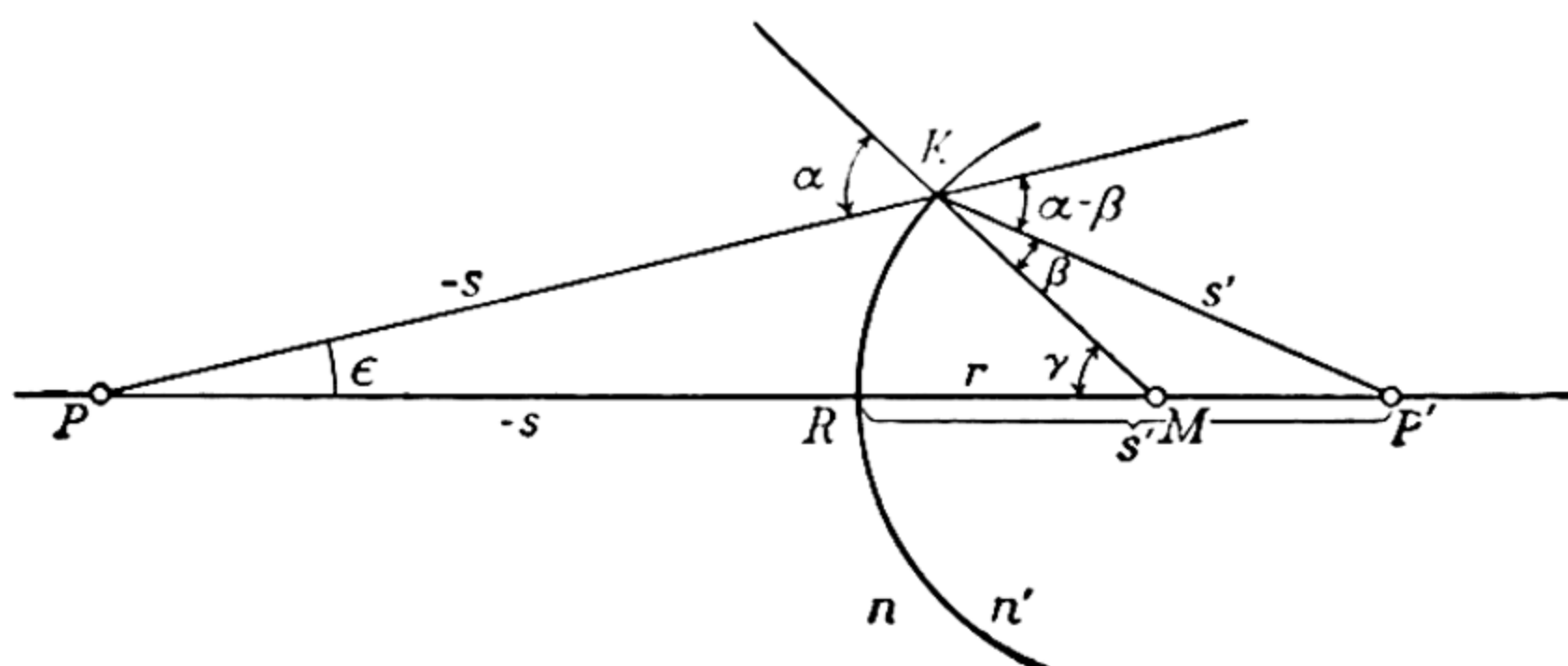


Fig. 6

the radius to the point  $K$  and denote the angle  $RMK$  by  $\gamma$ . Then, using the sine law of trigonometry and replacing the sines by the angles themselves, we have (see figure)

$$\frac{r - s}{-s} = \frac{\alpha}{\gamma} \quad \text{and} \quad \frac{s' - r}{s'} = \frac{\beta}{\gamma}. \quad \dots \dots (17)$$

Dividing one equation by the other and using the law of refraction in the form  $na = n'\beta$ , we have

$$\frac{(s - r)s'}{(s' - r)s} = \frac{n'}{n}, \quad \dots \dots (18)$$

or 
$$n \left( \frac{1}{r} - \frac{1}{s} \right) = n' \left( \frac{1}{r} - \frac{1}{s'} \right). \quad \dots \dots (19)$$

The angles  $\alpha$  and  $\beta$  are not present in the result. The value of the expression

$$n_i \left( \frac{1}{r} - \frac{1}{s_i} \right)$$

remains constant, and it is said to be a *surface invariant*. Equation (19) may be written in the form

$$\frac{-n}{s} + \frac{n'}{s'} = \frac{n' - n}{r}. \quad \dots \dots (20)$$

Comparing with equation (7), p. 399, and putting  $e = s$  and  $e' = s'$  yields \*

$$f = -\frac{nr}{n' - n}, \quad f' = \frac{n'r}{n' - n}, \quad \cdot \cdot \cdot \cdot \quad (21)$$

and hence

$$\frac{f'}{f} = -\frac{n'}{n}. \quad \cdot \cdot \cdot \cdot \quad (22)$$

If (22) is substituted into (11) and if  $u$  is substituted for  $\tan u$ , the result is the important Helmholtz-Lagrange equation

$$nyu = n'y'u'. \quad \cdot \cdot \cdot \cdot \quad (23)$$

Thus the expression  $n_i y_i u_i$  is also an invariant. Equation (23) represents the relation between angles and lateral distances in the object and image spaces even when there are many refracting surfaces, whereby it should be noticed that a large image corresponds to a narrow pencil and vice versa.

With the help of (22) we can now determine the combined effect of two successive imaging processes in the general case where the indices in object and image spaces are not the same. Combining equation (14), p. 402, with  $d = D + f_1' - f_2$  and neglecting  $d$ , we have

$$\frac{1}{f'} = \frac{f_1' - f_2}{f_1' f_2'} = \frac{1}{f_2'} - \frac{f_2}{f_1' f_2'} = \frac{1}{f_2'} + \frac{n_2}{n_2' f_1'}, \quad \cdot \cdot \quad (24)$$

or, introducing the power  $\phi$ ,

$$n_2' \phi' = n_2 \phi_1' + n_2' \phi_2'. \quad \cdot \cdot \cdot \cdot \quad (25)$$

Thus for a lens of index  $n_1$  having a medium of index  $n_0$  to the left and one of index  $n_2$  to the right, the last equation becomes

$$n_2 \phi' = n_1 \phi_1' + n_2 \phi_2',$$

where  $\phi_1'$  and  $\phi_2'$  are, respectively, the powers of the front and rear surfaces.

Up to now it has been shown that refraction by a spherical surface can be described by a relation of the form of equation (7) by identifying  $-nr/(n' - n)$  with  $f$  and  $n'r/(n' - n)$  with  $f'$ , and by measuring object and image distances from the plane normal to the axis at the vertex of the sphere. In this case the focal planes are really distant  $f$  and  $f'$ , respectively, from this plane, which thus represents the coincident pair of principal planes. Thus, putting  $s = -\infty$  in (20) and combining with (21), we get  $s' = f'$ ; in a similar way, for  $s' = \infty$  we get  $s = -f$ . If  $x$  and  $x'$  are now measured from the focal planes ( $s = x + f$  and  $-s' = -x' - f'$ ), then the first collinear equation of (6)

\* The factors  $1/n$  and  $1/n'$  imply only a difference of scale in the two spaces.

results. It remains to show that the second one is also satisfied. In order to find the object-image relation for a point  $Q$  near the axis (fig. 7), imagine the line  $PP'$  together with the other rays to be rotated through a small angle about the centre  $M$  of the sphere. Then  $P$  goes to  $Q$  and  $P'$  goes to  $Q'$ . The arcs  $PQ$  and  $P'Q'$  may, to the approxi-

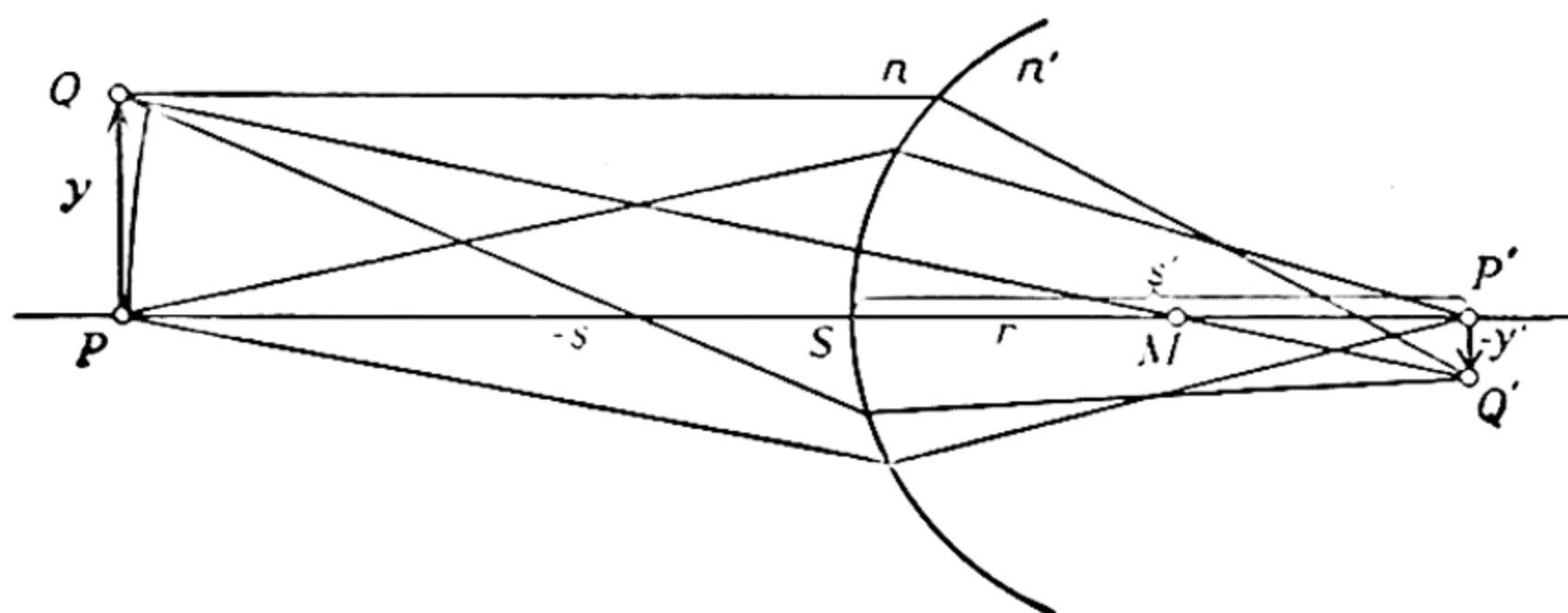


Fig. 7

mation adopted, take the place of the axial distances  $y$  and  $y'$ . Then, from the figure and from equation (19),

$$\frac{y'}{y} = \frac{s' - r}{r - s} = \frac{s' \left( \frac{1}{r} - \frac{1}{s'} \right)}{s \left( \frac{1}{s} - \frac{1}{r} \right)} = \frac{s' n}{s n'} = -\frac{s' f}{s f'}. \quad (26)$$

If  $x$  and  $x'$  are measured from the focal planes, so that we take  $-s = -x - f$  and  $s' = x' + f'$ , then the use of  $xx' = ff'$  gives us

$$\frac{y'}{y} = -\frac{x'f + ff'}{xf' + ff'} = -\frac{x'(f + x)}{f'(f + x)} = -\frac{x'}{f'} = -\frac{f}{x}, \quad (27)$$

so that also the second collinear equation is fulfilled.

(b) *Imaging of points near the axis by means of wide pencils.*

As soon as wider pencils of rays are used, the sine and the angle may no longer be taken equal and it is found that the image distance  $s'$  depends on the angle  $\epsilon$ . Rays emanating from  $P$  no longer re-unite in axial point  $P'$  but enclose a curve (or, in space, a surface of rotation) called a *caustic*. It is possible, however, to find mathematically a refracting surface such that all rays coming from  $P$  will converge again at  $P'$ . To obtain the analytical equation it is only necessary to apply the condition of constancy of optical path. For the case of reflection this surface is obviously an ellipsoid of rotation having the given points as foci, and it passes over into a paraboloid of rotation as  $P$  is moved to infinity.

The imaging of an axis point is of little practical importance, for



even the neighbouring parts of an object lying in a plane normal to the axis are not sharply focused. If the problem is to form an image not only of an axial point but of an entire element of area lying normal to the axis—something that can be done by using several refracting surfaces—then a condition first discovered by Abbe must be satisfied. This condition will now be deduced, using Fermat's principle (p. 395). Let the optical system, say a thick lens, be represented by the curved lines in fig. 8. The point  $P$  has its image at  $P'$ ; and  $P_1$ , which is in a plane through  $P$  normal to the axis, has its image at  $P_1'$ , which lies in a plane through  $P'$  and normal to the axis. Rays drawn through  $P$  and  $P_1$  parallel to the axis cross at  $F'$ , while parallel rays drawn through this pair of points at an angle  $u$  to the axis will meet at some point  $F_1'$ . It will not be concluded that  $F'$  and  $F_1'$  lie in a plane normal to the axis; for this amounts to assuming that an infinitely distant plane is imaged, point for point, in a focal plane. This is more than we wish

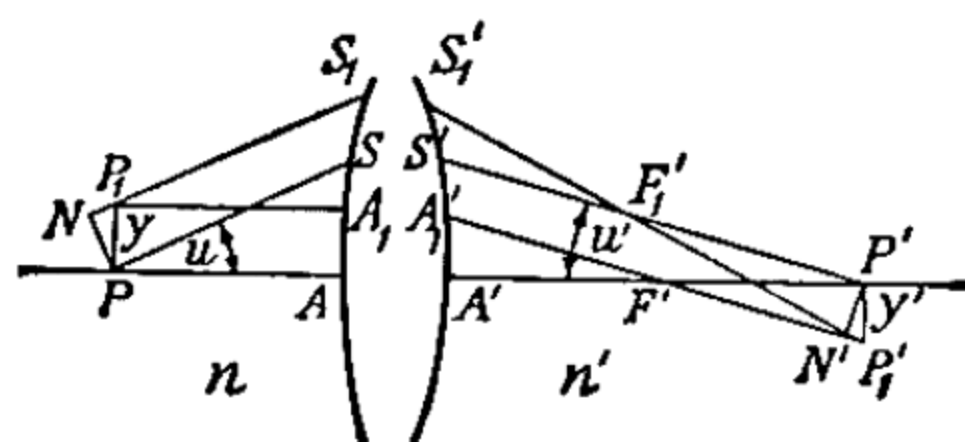


Fig. 8

to prove, since our deduction is limited to an *element* of area. Let the parallels to the axis enter the optical system at  $A$  and  $A_1$  and leave it at  $A'$  and  $A_1'$ . Similarly, let the oblique parallels enter at  $S$  and  $S_1$  and leave at  $S'$  and  $S_1'$ . According to the Fermat principle, the following relations will hold for the *optical* paths, which are denoted by parentheses:

$$(PAF'P') = (PSF_1'P') \quad \text{and} \quad (P_1A_1F'P_1') = (P_1S_1F_1'P_1'). \quad (a)$$

The paths  $(P_1A_1A_1'F')$  and  $(PAA'F')$  are equal, since  $F'$  is the intersection of two rays that are normal to an orthogonal surface (cf. p. 402). Further,  $(F'P_1')$  and  $(F'P')$  are equal as far as small quantities of second order, because of the smallness of  $(P'P_1')$ . Neglecting these term, we have

$$(PAF'P') = (P_1A_1F'P_1'). \quad . \quad . \quad . \quad . \quad . \quad (b)$$

$$\text{Combining with (a),} \quad (PSF_1'P') = (P_1S_1F_1'P_1'). \quad . \quad . \quad . \quad . \quad . \quad (c)$$

Drop a perpendicular from  $P$  on to  $P_1S_1$  and call the foot of it  $N$ . Then, because both rays are normal to the same orthogonal surface,

$$(PSF_1') = (NS_1F_1'). \quad . \quad . \quad . \quad . \quad . \quad (d)$$

Subtraction of (d) from (c) yields

$$(F_1'P') = -(NP_1) + (F_1'P_1'). \quad . . . . (e)$$

Now erect a perpendicular  $P'N$  to the line  $F_1'P'$ . As far as second order quantities,

$$(F_1'P_1') - (F_1P') = (N'P_1'), \quad . . . . (f)$$

that is,

$$(NP_1) = N'P_1'. \quad . . . . (g)$$

If the index of refraction is  $n$  in the object space and  $n'$  in the image space (the two indices would be different, for instance, for an immersion microscope objective), then

$$(NP_1) = ny \sin u, \quad (N'P_1') = n'y' \sin u'; \quad . . (h)$$

whence, according to (g),

$$ny \sin u = n'y' \sin u'. \quad . . . . (28)$$

In place of the Helmholtz-Lagrange equation we have the relation (28), while according to (11) and (22) collinear projection amounts to the same relation with the tangents in place of the sines of the angles. This shows that collinear projection of the entire space, using arbitrarily wide pencils of rays, is fundamentally impossible. Similarly, we could proceed to show, as in the deduction of the sine law, that if the sine condition holds for a given pair of axial points it is impossible for it to be valid for any other pair. The special case of plane reflection is an exception. Here the whole of space may be imaged collinearly using wide pencils, for since  $n = n'$ ,  $y = y'$  and  $u = -u'$ , the sine and tangent conditions are compatible in this instance.

### (c) *Imaging of points distant from the axis by means of narrow pencils*

Even an *elementary* pencil from a point off the axis will fail to focus at a point after refraction at a spherical surface. Moreover, it is readily seen that rays in the meridian plane—the plane containing the axis and the point—have a different point of intersection from those in the plane normal to this one. Nevertheless, according to the theorem of Malus, every elementary pencil has an orthogonal surface.

Imagine the pencil so limited that the orthogonal surface in the image space is bounded by two pairs of neighbouring lines of curvature (fig. 9). As we know, neighbouring normals to the surface intersect along the lines of curvature. The normals at  $A$  and  $B$  intersect at  $E$ , those at  $C$  and  $D$  have their point of intersection at  $F$ . The straight line  $EF$  is normal to the principal ray drawn through the

centre of the element of surface. In the same way, the second pair of lines of curvature furnishes a line of intersection  $GH$  which is in a direction at right angles to  $EF$ . The distances  $AE$  and  $AG$  represent principal radii of curvature of the surface element. The rays reunite at a single point only when these two radii are equal, otherwise there are two focal lines whose distance apart is called the *astigmatic difference*.

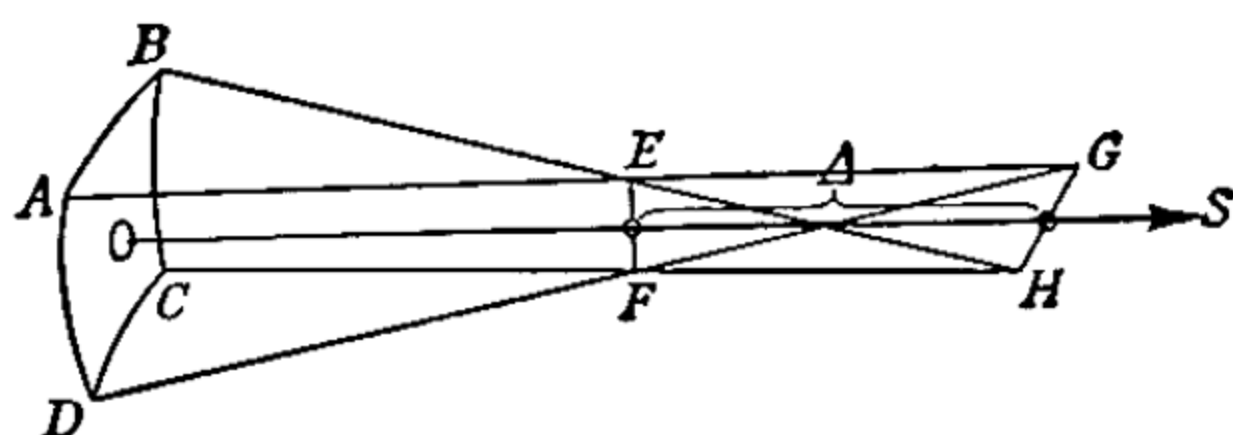


Fig. 9

*ference*  $\Delta$ . An astigmatism of this kind occurs when the principal rays meet a spherical refracting surface obliquely.

*Ex. 104.* Show that the laws of reflection and refraction hold also for the angles which the incident and reflected (or refracted) rays make with any plane containing the normal to the surface of separation.

*Ex. 105.* In a prism spectrograph, why are the lines curved?

*Ex. 106.* Compute the focal lengths and locate the principal planes of a thick lens of index  $n$ .

*Ex. 107.* Find the form taken by the sine law for an infinitely distant axial point object. (Introduce the height  $h$  at which a ray coming from an axial point cuts the first surface of the system.)

#### 4. The Resolving Power of Optical Systems.

Even if the science of geometrical optics would enable us to construct an objective which, by compensation of the various imperfections of single-point image formation, brought all rays from a point source (e.g. a distant star) to a point focus, observation would still show the image to be a small disc surrounded by a system of rings. This pattern is due to diffraction by the edges of the lenses and diaphragms. Although the point of observation is in the finite region—in our example the focal plane of the telescope objective—the case is one of Fraunhofer diffraction, since diffraction at the aperture gives rays in various directions; this is equivalent to infinitely distant point sources of various intensities. The evaluation of the diffraction integral may be performed for a circular opening with the aid of Bessel functions. Qualitatively, the intensity variation is not different from that found for a slit. However, for normal incidence, the first minimum is in a direction  $\theta$  given by

$$r \sin \theta = 0.61\lambda, \quad . . . . . (29)$$



where  $r$  is the radius of the opening. A second point can be separated from the first if the angular distance between the two is at least so great that the maximum for the second point falls upon the first minimum in the image of the other. Thus, for a circular aperture,

$$\Delta\theta = \frac{0.61\lambda}{r}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (30)$$

The "limit of resolution" thus defined cannot be influenced by increasing the magnification; its value, however, is inversely proportional to  $r$ , and this is the reason for the large dimensions of telescope objectives.

*Ex. 108.* Calculate the resolving power of a prism in terms of its dimensions and the dispersive power  $dn/d\lambda$  of the substance of which it is made.

## 5. The Fundamentals of Interference Optics. Interference Fringes.

In the construction of a certain class of optical instruments, the so-called "interference" instruments, the wave nature of light is taken into consideration in so far as several coherent pencils are combined, this being done in such way as to allow for their dif-

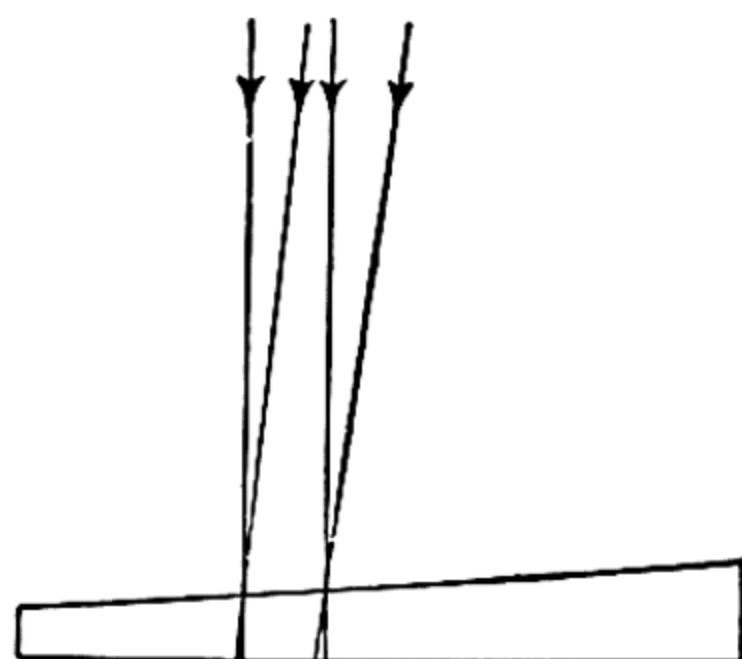


Fig. 10

ference in phase. The diffraction phenomena occurring at the boundaries are not of essential importance in this type of device. If a plane wave be reflected from a wedge of small angle (fig. 10), the superposition of the waves reflected from the front and back sides of the wedge will yield an increased or decreased resulting disturbance, depending upon the phase difference of the emerging waves. But the phase difference depends only upon the optical path between the two faces of the wedge; thus there will be places where the disturbance is intensified and others where it is weakened. If the incident beam is exactly parallel we shall not be able to see these places at all. As a matter of fact, it is physically impossible to produce such a beam, since it would require a source consisting of a single point of light (whose surface brightness would therefore have to be infinite) placed

at the focus of a lens. In reality a source of light always has a finite extension, and since every point of the focal plane corresponds to a given parallel bundle of rays, the resultant beam always has a certain divergence. We shall assume, however, that this divergence is so small and that the *thickness* of the wedge is also so small that the path difference at a given point on the wedge is the same for all rays. The rays incident at this point then give a resulting illumination whose intensity depends upon whether the path difference at the front surface is an integral number of whole wave-lengths or half wave-lengths. In the former case, the resulting illumination is intense; in the latter case, weak. The eye or a photographic objective focussed on the upper surface of the wedge will observe a system of light and dark bands, or "interference fringes". These are the "curves of equal

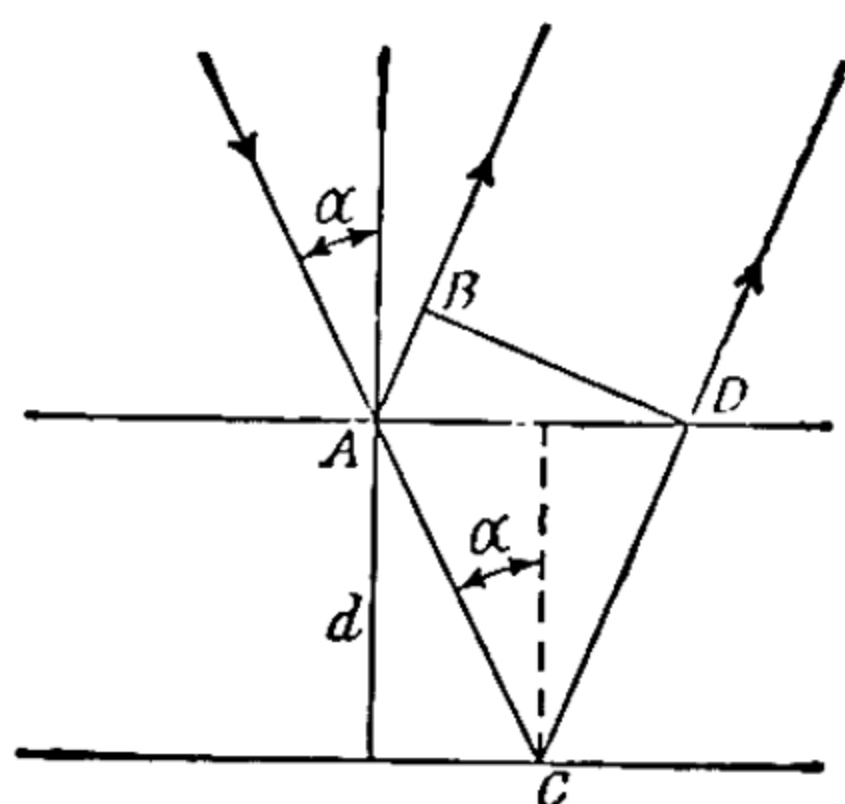


Fig. 11

thickness" of the wedge. As the angle of the wedge is increased, or as its thickness is made greater, even the rays at a given point have different path differences and the visibility of the fringes decreases considerably even for monochromatic light.

On the other hand, if we allow a divergent beam to fall upon a plane parallel plate, the difference in inclination gives rise to a difference in path of the rays reflected from the two surfaces. If the index of refraction of the plate or film is the same as that of the surrounding medium, as in the case of the Fabry-Pérot interferometer, the path difference for an angle of incidence  $\alpha$  is (cf. fig. 11)

$$AC + CD - AB = \frac{2d}{\cos \alpha} - 2d \tan \alpha \sin \alpha = 2d \cos \alpha. \quad (31)$$

In those directions for which  $2d \cos \alpha = m\lambda$  there is light, while for directions given by  $2d \cos \alpha = (2m + 1)\lambda/2$  there is darkness. All rays making an angle  $\alpha$  with the normal to the surface yield a circle in the focal plane of an objective focussed for infinity, on account of the com-

plete symmetry about the normal, so that the fringe system consists of concentric circles. The rings become narrower if the thickness of the plate is increased, but—unlike the previous case—their visibility is limited only to the degree that the light source is not strictly monochromatic. If we assume that a spectral line consists of, say, two strictly monochromatic components, there will be a condition where the interference fringes of one component fall between those of the second, making the resultant pattern indistinct. In reality, no spectral line is strictly monochromatic; the fact that a radiating atom sends forth a succession of finite wave trains means that the spectral line actually consists of a continuous distribution having a sharp maximum. This may be seen by considering the finite wave train analysed into an equivalent sum of Fourier components. The result is that the interference effects due to the different wave-length regions overlap when the path difference is large.

If we have but two interfering rays, the vector diagram indicates a decline of intensity according to the formula

$$I = I_{\max.} \cos^2 \frac{\phi}{2}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (32)$$

where  $\phi$  is the phase difference. If we allow a larger number of rays to interfere we obtain a considerably steeper decline of intensity, and hence greater resolving power for wave-lengths which are close together (cf. p. 52). This condition may be realized by lightly silvering the two surfaces of a plane parallel plate, so that these surfaces still transmit part of the light falling on them. Light transmitted by the system consists of a large number of partial rays (in reality infinitely many, with decreasing intensity) which have experienced various numbers of reflections. This arrangement is used as a spectroscope of very high resolving power, and is the familiar interferometer of Fabry and Pérot. If we had a finite number of partial beams of equal intensity we could apply the results obtained for the grating by means of the vector diagram. We found the resolving power of a grating to be (p. 384)

$$R = Nm. \quad . \quad . \quad . \quad . \quad . \quad . \quad (33)$$

The difference here is that the number of rays of a given intensity is small, while the order—i.e. the path difference for neighbouring rays—is large in the present case. If the thickness of the plate is 0.25 cm. and we have green light,  $\lambda = 5 \times 10^{-5}$  cm. incident normally,  $m = 2d/\lambda = 10^4$ , while  $m$  is usually less than 5 for the grating. The fact that we really have infinitely many rays whose intensity decreases geometrically modifies the considerations somewhat. There is no longer a sidewise minimum, and we must impose a criterion



which determines the distance of a second maximum which just may be perceived separately. In practice, an interferometer of this type is characterized by the equivalent number of rays of equal intensity—a number around 20. Thus in our case this implies a resolving power of about  $2 \times 10^5$ . To obtain an idea of what this means, let us consider the two D-lines of sodium,  $\lambda = 5890, 5896$ . The fractional wave-length difference is about  $10^{-3}$ , so that the above interferometer would be capable of resolving two lines whose separation is about one two-hundredth that of the sodium lines.

One disadvantage of this type of interference spectroscopy, however, is its small region of dispersion, which makes it necessary to use this instrument by crossing it with another type of dispersive instrument, e.g. a grating. If the incident light embraces a considerable wave-length region, the resulting interference pattern will be hopelessly complicated. The “region of dispersion”  $\Delta\lambda$  is that difference in wave-length for which the  $m$ -order interference pattern of one wave-length coincides with the  $(m + 1)$  order pattern of the other. Then

$$m(\lambda + \Delta\lambda) = (m + 1)\lambda$$

or 
$$\Delta\lambda = \frac{\lambda}{m} \quad . \quad . \quad . \quad . \quad . \quad . \quad (34)$$

Thus the region of dispersion in the example above is one-tenth the separation of the sodium lines.

*Ex. 109.* Calculate the difference in path of rays reflected from the two surfaces of a plane parallel plate of glass if the rays are parallel and enter the plate at an angle of incidence  $\alpha$ . The outside medium is air.



## PART IV

### THE THEORY OF ELECTRICITY

#### II. THE ATOMISTIC NATURE OF ELECTRICAL PHENOMENA

In the preceding part of this work we considered a theory which found its complete formulation, for stationary bodies, in the Maxwell Field Equations. This theory made use of a number of experimentally determinable material constants—the conductivity  $\sigma$ , the dielectric constant  $K$  and the magnetic permeability  $\mu$ , as well as the index of refraction  $n$  which is derivable from these quantities. But, as mentioned in several instances, these quantities are not invariable material constants, but depend upon temperature, pressure, and—for rapidly varying fields—upon the frequency. By its very nature, the field theory is powerless to give any information concerning this variability of  $\sigma$ ,  $K$  and  $\mu$ ; rather, we must try to construct a picture, based upon suitable assumptions as to the structure of matter and of electrical charge, which furnishes the correct functional dependence of these quantities upon the several variables. Thus in contrast with the field theory, which uses only macroscopic quantities, we must employ concepts of a microscopic nature which, in part, are not capable of direct verification and whose validity can be tested only by examining the conclusions obtained from their use.





## CHAPTER XXII

### ELECTROLYTIC CONDUCTION

#### 1. The Fundamental Phenomena of Electrolytic Conduction and their Interpretation.

We begin with the consideration of a phenomenon the experimental laws of which are not explicable by the Maxwell equations, but have been known for a long time. These empirical relationships led, at an early period, to hypotheses concerning the nature of electricity. The phenomenon referred to is the conduction of electricity in solutions of acids, bases, or salts—the so-called electrolytes.

*Electrolytic conduction follows Ohm's Law at all field strengths from the weakest to the strongest, i.e. at constant temperature the conductivity is independent of the field strength.* Only recently have small departures from Ohm's Law been found for the highest fields. This type of conduction is accompanied by a transport of matter. The constituents of the substance in solution are deposited at the electrodes in accordance with the familiar Faraday Law, which states that *the electrolytic liberation of one chemical equivalent\* of a substance always requires the same quantity of electricity, viz.:*

$$F = 2.893 \times 10^{14} \text{ e.s.u. gm. sec./mol} = 9649 \times 10^4 \text{ amp. sec./kg. mol.}$$

A chemical equivalent is understood to mean that mass of an element or group of elements (chemical radical) which could displace one gramme molecule (1 mol) of atomic hydrogen, i.e. 107.9 gm. of univalent silver, (1/2)(63.6) gm. of bivalent copper, (1/3)(138.9) gm. of trivalent lanthanum, &c.

If we look upon the atomic nature of matter as a proven fact, as we now have every reason to do, the Faraday Law makes the atomic character of electricity seem extremely *probable*. One equivalent of silver contains  $L = 6.02 \times 10^{23}$  silver atoms. The transfer of this number of atoms is associated with the transfer of a quantity of electricity  $F$ , so that *on the average* each atom conveys a charge  $e$  given by

$$e = \frac{F}{L} = \frac{9650 \times 10^4}{6.02 \times 10^{26}} = 1.60 \times 10^{-19} \text{ coulombs} = 4.80 \times 10^{-10} \text{ e.s.u.}$$

Since we always find this value in every experiment, it is natural to

\* In the M.K.S. system the kilogramme mol must be used.

assume that this quantity is not only the *average* value—which is all we are justified in concluding thus far—but that every univalent atom actually carries a charge  $e$ , every bivalent atom a charge  $F/(L/2) = 2e$ , &c. The charge  $e$  may be called the elementary unit of electrical charge—the “atom” of electricity. For reasons which we shall discuss later, it is most commonly called the charge upon the electron. The existence of this element of charge has been demonstrated beyond question by a large number of experiments. On account of their behaviour in electrolytes, we call atoms or atom groups which carry one or more elementary charges *ions* (Greek *ἰόν*, going). What is the origin of the charged particles? We now know from X-ray crystallography that a simple crystal of a salt, e.g. an NaCl crystal, is built up not of neutral atoms, but of positive  $\text{Na}^+$  ions and negative  $\text{Cl}^-$  ions. When dissolved in water the lattice structure is destroyed, but it is very probable that the ions retain their charges. It is, of course, possible that a certain part unite to form neutral NaCl molecules.

In the case of electrical discharge in gases, the carriers of the current are formed by the field itself, as we shall see later. This cannot be true in the present case, since we find the same value of the conductivity for the weakest fields as we do for very strong ones.

If, in one cubic centimetre of an electrolyte, we have  $N_+$  positive ions of valence  $z_+$ , and  $N_-$  negative ions of a substance of valence  $z_-$ , and if the speeds of the two kinds of ions are  $v_+$  and  $v_-$  respectively, then the charge flowing through unit cross-section per second is

$$i = (z_+ N_+ v_+ + z_- N_- v_-) e, \quad . \quad . \quad . \quad . \quad . \quad (1)$$

for as many positive ions pass through the unit cross-section each second as are contained in a prism of length  $v_+$ , i.e.  $N_+ v_+$ . Similarly,  $N_- v_-$  ions pass in the opposite direction; this corresponds to a stream of positive charges *in* the direction of the field. Thus  $i$  is the value of the current density. Since the solution as a whole is neutral,

$$N_+ z_+ = N_- z_-, \quad . \quad . \quad . \quad . \quad . \quad (2)$$

so that

$$i = N_+ z_+ e (v_+ + v_-). \quad . \quad . \quad . \quad . \quad . \quad (1')$$

If Ohm's Law is to hold, i.e. if  $i$  is to be proportional to the field strength  $E$ , then according to equation (1),  $v_+$  as well as  $v_-$  must be proportional to  $E$  and independent of the time. One might at first imagine that a constant force  $zeE$ , which acts on an ion of valence  $z$ , would produce an acceleration, i.e. a velocity varying with time. But in addition to the field, the ion is subject to the frictional resistance of its motion through the solvent. If we make the somewhat bold assumption that Stokes's Law (p. 222) holds even for spheres



of atomic size, we have a resistance  $f$  proportional to the speed:

$$f = 6\pi\eta Rv, \quad . . . . . (3)$$

where  $R$  is the radius of the sphere and  $\eta$  is the viscosity of the fluid. After a short period of acceleration there is equilibrium between the force exerted by the field and the resisting force, so that

$$v = \frac{zeE}{6\pi\eta R}. \quad . . . . . (4)$$

The velocity attained in unit field is called the *mobility* of the ion:

$$u = \frac{v}{E} = \frac{ze}{6\pi\eta R}. \quad . . . . . (5)$$

The mobility is usually referred to a field of 1 volt/cm. Introducing the mobility into equation (1) (p. 418), we have

$$i = (z_+N_+u_+ + z_-N_-u_-)eE = N_+z_+eE(u_+ + u_-), \quad . (6)$$

and the conductivity becomes

$$\sigma = (z_+N_+u_+ + z_-N_-u_-)e = N_+z_+e(u_+ + u_-). \quad . (7)$$

Thus the value of the conductivity alone does not give the mobility of each kind of ion, but only the sum of the two. In addition, we must know another relation between the two mobilities, e.g. that they are equal, or that their ratio has a given value. This may be determined experimentally (cf. below). We see also from equation (5) that the mobility is greater the smaller the viscosity of the solvent. Now the latter decreases, for all liquids, with rising temperature, so that we have an explanation of the increase of electrolytic conductivity with increasing temperature.

The ratio of the mobilities may be determined by measuring the changes in concentration occurring at the electrodes when a current passes. The natural diffusion tending to cancel these differences in concentration is prevented by inserting a porous diaphragm into the vessel. We shall compute the mobility ratio in this way for an electrolyte whose ions are both univalent. There are deposited at the cathode not only the cations which have migrated there, but also those set free by the moving away of their oppositely-charged partners. That is, only *one* kind of ion need have appreciable mobility in order that there may be deposits at both electrodes. If it were otherwise, the electrolyte would have an excess charge at the electrodes; this charge would equalize itself by deposition of the superfluous ions. Fig. 1 is a schematic diagram of an electrolytic cell in which we assume the ratio of mobility of the cation to that of the anion to be 3 : 2. Only

those cations which are set free by the departure of the anions have any effect on the *change in concentration* at the cathode. Hence the

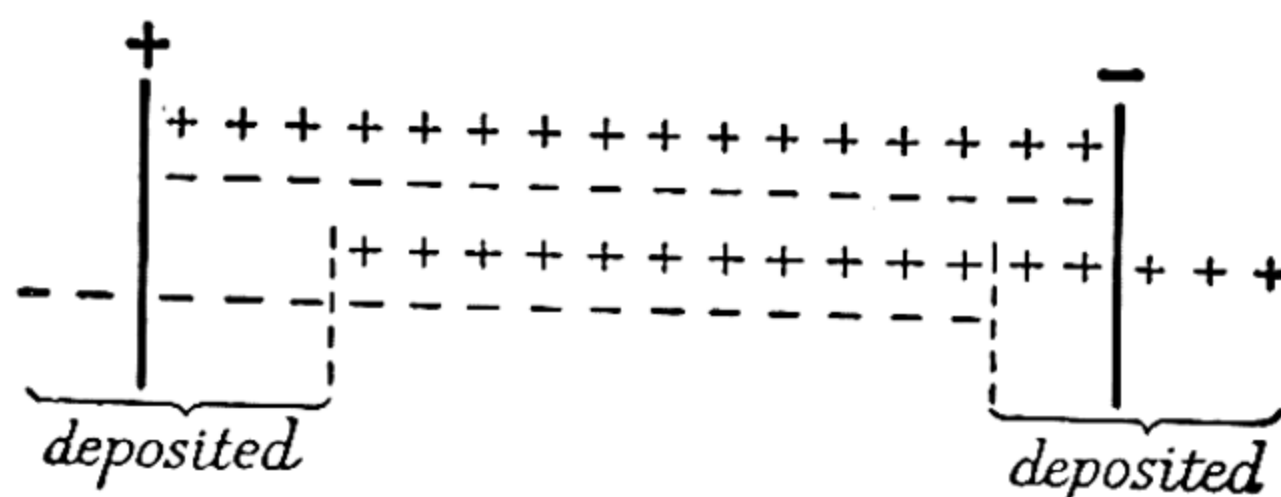


Fig. 1

change in concentration at the cathode, expressed in gramme molecules of the solute is

$$\Delta n_K = \frac{N_- u_- E}{L}, \quad . . . . . (8)$$

where  $E$  is the field strength and  $L$  is Avogadro's Number—the number of molecules per mol. The change in concentration at the anode is

$$\Delta n_A = \frac{N_+ u_+ E}{L}. \quad . . . . . (9)$$

Since  $N_- = N_+$  when both ions are univalent, we have

$$\frac{\Delta n_K}{\Delta n_A} = \frac{u_-}{u_+}. \quad . . . . . (10)$$

*The losses in concentration at the respective electrodes, expressed in gramme molecules of the solute, are inversely proportional to the mobilities of the ions arriving there.* This relation was given by Hittorf.

In this form, the law holds also for an electrolyte whose ions have valences  $z_+$  and  $z_-$ . The chemical formula of the solute is then  $K_{z_-} A_{z_+}$ , where  $K$  represents the cation,  $A$  represents the anion, and  $z_+$  and  $z_-$  are relatively prime. It is true that we now have  $z_- N_- = z_+ N_+$  in place of simply  $N_- = N_+$ , but the decrease in concentration at the anode, expressed in gramme molecules of the solute, is calculated from the decrease in the number of cations by dividing by  $z_- L$  (cf. the chemical formula), so that the valences cancel in forming the ratio.

The ratio of mobilities gives also the ratio of the parts of the total charge carried by each kind of ion. Denoting the fraction of the total charge carried by the cations by  $q_+$ , that carried by the anions by  $q_- = 1 - q_+$ , we have by equation (6) (p. 419),

$$\begin{aligned} q_+ i &= z_+ N_+ u_+ e E, \\ q_- i &= z_- N_- u_- e E, \end{aligned}$$

whence 
$$\frac{q_+}{q_-} = \frac{q_+}{1 - q_+} = \frac{u_+}{u_-}. \quad . . . . . (11)$$

Following Hittorf,  $q_+$  and  $q_-$  are called the *transport numbers*.

According to equation (6) (p. 419) and equation (11) immediately above, the individual mobilities of the ions may be determined from measurements of conductivity and transport ratios. These mobilities are very small. The following are several values in cm. sec.<sup>-1</sup>/volt cm.<sup>-1</sup>:

Li <sup>+</sup>	$36.8 \times 10^{-5}$
Na <sup>+</sup>	$46.0 \times 10^{-5}$
K <sup>+</sup>	$67.6 \times 10^{-5}$
Mg <sup>++</sup>	$47.6 \times 10^{-5}$ .

If we use these values in conjunction with Stokes's Law to determine the ionic radii, we obtain values of the correct order of magnitude,  $10^{-8}$  cm. But we see that these radii are not the same as those of the free ions in the gaseous state, for—contrary to well-founded results of atomic physics—they decrease in the order Li-Na-K, rather than increase. The reason is that there exists a stronger electrostatic field at the "surface" of a small ion than at the surface of a large one, assuming the charge to be symmetrically distributed over a sphere. This field causes the attachment of molecules of the solvent, the effect becoming more feeble toward the outside. The observed radius is, then, the effective radius of the ion together with its shell of molecules of the solvent. For this reason the mobility of the bivalent Mg<sup>++</sup> ion is not very much different from that of the similar univalent Na<sup>+</sup> ion, although for a given field, the force on the former is twice as great. If we omit from consideration the constituent ions of water, H<sup>+</sup> and OH<sup>-</sup>,\* we may say that the mobilities of all ions are approximately equal, for if an ion has a greater charge, and hence a greater driving force, it will also have a greater effective radius, and hence experience greater resistance to its motion.

## 2. Dependence of Electrolytic Conductivity upon Concentration. The Theory of Debye-Hückel and Onsager.

Let there be given an electrolyte in which  $p$  gramme equivalents are dissolved in each cubic centimetre of solution, and let the fraction ionized be  $\alpha$ , while the fraction  $1 - \alpha$  remains in the form of undissociated molecules, and so has no effect on the transport of electricity.  $\alpha$  is called the degree of dissociation. The conductivity is then given by equation (7) (p. 419) as

$$\sigma = \alpha p L e (u_+ + u_-) = \alpha p F (u_+ + u_-). \quad (12)$$

The ratio  $\sigma/p$ , the *equivalent conductivity*  $\Lambda$ , would be independent of the concentration if

1.  $\alpha$  were independent of the concentration, and if
2. the mobilities  $u_+$  and  $u_-$  were also independent of this quantity.

\* These ions have abnormally large mobilities:  $329 \times 10^{-5}$  for H<sup>+</sup> and  $180 \times 10^{-5}$  for OH<sup>-</sup>. There are undoubtedly other factors which play a rôle here.



The observed very pronounced dependence of  $\Lambda$  upon the concentration can be explained only by discarding one or both of the above assumptions. The first alternative—rejection of 1 and retention of 2—corresponds to the classical theory of electrolytes connected principally with the name of Arrhenius.\* At first sight the fundamental assumption of this theory seems very plausible: the frictional forces are assumed to result wholly from interaction with the molecules of the solvent. This is quite probable; for example, in a molar solution of NaCl the ratio of the number of ions per c.c. to the number of water molecules is about 18 : 1000. This hypothesis, however, overlooks the enormous electrostatic forces exerted by the elementary charge within a range of a few atomic radii.

The presence together of molecules and their products of dissociation represents a type of chemical equilibrium governed by a thermodynamic law known as the Law of Mass Action (cf. p. 542 *et seq.*). For the case at hand this rule takes the form

$$\frac{\alpha^2}{1-\alpha} p = K(T),$$

where the mass action constant  $K$  depends only upon the temperature  $T$ . From this we see that  $\alpha$  approaches the value unity as the concentration is decreased. For small values of the concentration we may set  $\alpha = 1$  in the numerator, in which case we see that the degree of association  $1 - \alpha$  is proportional to the equivalent concentration. If the classical formulation is correct, the ratio of the equivalent conductivity  $\Lambda$  for finite concentration to that extrapolated to zero concentration,  $\Lambda_0$ , must be equal to the degree of dissociation  $\alpha$ , which in turn must follow the law of mass action. This is found to be true for the greater number of weak electrolytes, i.e. substances—chiefly organic acids and salts—having low values of  $\Lambda$  at moderate concentrations. On the other hand, for strong electrolytes—the commoner inorganic acids, bases and salts—we find great deviations. In some instances the expression

$$\frac{\alpha^2}{1-\alpha} p = \frac{\left(\frac{\Lambda}{\Lambda_0}\right)^2}{1 - \frac{\Lambda}{\Lambda_0}} p$$

varies by several powers of ten. For this reason theories based on the alternative point of view were proposed from the very beginning. These speculations were based upon the assumption that dissociation is complete even for relatively concentrated solutions (up to perhaps 1/100 molar), but that the ions exert mutually hindering forces upon

\* S. Arrhenius, *Zeitschr. phys. Chem.*, **1**, p. 631 (1887). *Lehrbuch der Elektrochemie*, Leipzig, Quandt & Händel (1901).

each other. It was Debye and Hückel who first succeeded in overcoming the extreme mathematical difficulties attendant upon the computation of these inter-ionic forces. Their theory was considerably improved later by L. Onsager. We are able to give only the main features of the calculation here; the original papers should be consulted for details.\*

If we consider a given positive ion and note the number of ions of both kinds in its neighbourhood, there will be, when averaged over a long time or over many positive ions, a preponderance of negative ions in its immediate vicinity. If we compute the average distribution of charge in this region, we find it to consist of a spherically symmetric negative charge whose density falls off rapidly with distance from the centre. Similarly, a negative ion is surrounded by a cloud of positive charges. The manner in which these clouds of charge come about may be seen by considering a crystal lattice (fig. 1, p. 573). Here also, the nearest neighbours of each ion are ions of opposite sign. We may picture the structure of an electrolyte in a similar way except that the lattice structure is replaced by one which is more "plastic". The computation is made possible by imagining the charge in the neighbourhood of a single ion to be replaced by the average continuous volume charge obtained by averaging over the actual ion configurations from instant to instant. The density of this average ion cloud is indicated in fig. 2, Plate I, by different degrees of shading.

Besides the Stokes resistance, there is a retarding force of the following nature: it requires a certain time for the cloud of ions to attain its spherically symmetric equilibrium configuration. If the central ion is moved forward a distance  $ds$  in the time  $dt$ , the surrounding charges tend to retain their former configuration. This means that there is a deficiency of charge of opposite sign on the forward side and an excess on the rearward side. The electrostatic forces then retard the ion.

In addition to this "force of relaxation", there is a further retarding effect, the electrophoretic force. According to p. 218, when a sphere moves through a viscous fluid, the velocity of the fluid at the surface of the sphere is the same as that of the sphere, and decreases outward. If, then, negative ions travelling in the opposite direction predominate in the neighbourhood of the positive ion, the resultant effect of the surrounding ions is to oppose the motion of the positive ion; otherwise stated, this ion moves through a region of the solvent which itself is flowing in the opposite direction. The results obtained by quantitative investigation of these ideas are:

1. The mobility of a given ion depends not only upon the nature of the ion and the solvent, but also upon the nature of the other ion.

\* P. Debye and G. Hückel, *Phys., Zeit.* 24, p. 305 (1923). L. Onsager, *ibid.* 27, p. 388 (1926).



The independence of ionic motion is valid only for infinite dilution. The numbers given in § 1 (p. 421) are limiting values of the mobilities at zero concentration.

2. The two retarding forces are, like the Stokes resistance, proportional to the velocities of the ions, to a first approximation. Thus Ohm's Law remains valid, although only to a first approximation for low field strengths. As a matter of fact, M. Wien found considerable departures from Ohm's Law for very strong fields ( $10^5$  volts/cm.). He observed an increase in conductivity with increasing field, in good agreement with the theory.\* In the limit of infinitely strong fields the added retarding forces vanish entirely, so that the remaining departure from the limiting value of the conductivity for infinite dilution is correctly ascribed to association.

3. For very dilute solutions the departures from the limiting value of the conductivity, i.e. the quantities  $1 - \Lambda/\Lambda_0$  are no longer proportional to the concentration itself, as in the law of mass action, but to the square root of the concentration. This rule was advanced as an empirical fact by Kohlrausch long before the theory was proposed.

4. For a number of equally concentrated solutions of the same solute in different solvents, the departures are greater for the solvents of smaller dielectric constant. To see that the inter-ionic forces must increase with decreasing dielectric constant we need only recall that the expression for the force between two point charges has  $K$  in the denominator.

The development of the cloud of ions about a given ion is imperfect for rapidly alternating fields as well as for very strong fields. As a result, the added retarding forces are absent when high-frequency alternating currents, corresponding to a wave-length of 10 m. or less, are used. Going to still shorter waves, the electrolytic conductivity stops altogether when the wave-length reaches a few centimetres, as has been known for a long time both experimentally and theoretically. The reason is that the ions, on account of their inertia, are no longer able to follow the alternations of the field. This provides us with the connexion with optics. In the optical range, solutions of simple salts like NaCl are to be considered insulators, as is shown by their transparency.

*Ex. 110.* In order to gain an idea of the magnitude of the electrostatic forces calculate (a) the attraction between two elementary charges  $e = 4.80 \times 10^{-10}$  e.s.u. which are at a distance of  $10^{-8}$  cm. apart; (b) the attraction between the positive and negative ions in 1 mol of rock salt which are separated and placed at a distance apart equal to the earth's diameter (12,700 km.); (c) the surface pressure necessary to hold together a sphere of 10 cm. radius into which all the ions of one sign are packed.

\* M. Wien, *Ann. d. Phys.*, 83, p. 327 (1927).



## CHAPTER XXIII

### THE CONDUCTION OF ELECTRICITY IN GASES

#### 1. Direct Determination of the Elementary Electrical Charge by the Millikan Oil Drop Method.

Before we treat the theory of electrical discharge in gases we wish to decide whether the value of the elementary charge on a univalent ion found in electrolysis is merely an average value (this is all we can infer so far) or the true value of a smallest unit of charge. The experimental proof of the correctness of the latter assumption was given by Millikan. The arrangement used by him was essentially that described on p. 263. The procedure is to observe the motion, in an electric field, of small test bodies carrying a few elementary charges. Millikan used droplets of oil formed by atomization. Their motion was observed in a microscope, the droplets being illuminated from the side. The droplets are in a vertical electric field. With the field off, the terminal velocity of fall  $v_0$  is determined by the equilibrium between gravity and the Stokes resistance to motion through the air. If we assume—as we may for oil drops—that the droplet is spherical and of radius  $a$ , and if we call the density of the oil  $\sigma$ , that of the air  $\sigma_A$ , and  $\eta$  the coefficient of viscosity of the air, then we have the equation

$$6\pi\eta av_0 = \frac{4\pi}{3} a^3 (\sigma - \sigma_A) g. \quad . . . . . (1)$$

From a measurement of  $v_0$  we obtain the radius of the droplet, and thus its mass. If we now apply an electric field  $E$  which, with gravity, urges the drop downward, and if we again determine the velocity  $v_1$ , we have

$$6\pi\eta av_1 = e_e E + \frac{4\pi}{3} a^3 (\sigma - \sigma_A) g. \quad . . . . . (2)$$

The charge on the drop is then given by

$$e_e = \frac{4\pi}{3} a^3 \frac{(\sigma - \sigma_A) g \left( \frac{v_1}{v_0} - 1 \right)}{E}. \quad . . . . . (3)$$

The result of Millikan's experiment is that the charges  $e_i$  are *small*\* integral multiples of a unit whose magnitude is

$$e = 4.80 \times 10^{-10} \text{ e.s.u.} = 1.60 \times 10^{-19} \text{ coulombs.}$$

For the exact calculation it is necessary to introduce a kinetic theory correction to Stokes's Law, since the size of the drop is no longer large compared with the mean free path of the air molecules. Several investigators have expressed doubt as to the justification of some of the assumptions embodied in the experiment, viz. the accurately spherical form of the droplets, the identity of their density with that of the oil in bulk, and the decrease in mass due to evaporation during the period of observation. It is fairly certain, however, that any errors attributable to these causes are very small.

The Millikan experiment not only proves the existence of the quantum of electricity  $e$ , but also gives the most accurate value of this constant so far obtained. This quantity also furnishes a value of  $L$ , since the Faraday equivalent charge  $F = Le$  is directly measurable with great accuracy. These results

$$L = 6.02 \times 10^{23} \text{ mol}^{-1} = 6.02 \times 10^{26} \text{ (kg. mol)}^{-1}.$$

## 2. The Nature of the Cathode Ray Particles. The Electron.

If a glass tube provided with two electrodes is pumped down to a pressure of a few thousandths of a millimetre of mercury, and if an electrical potential of several thousand volts is applied, one observes a green fluorescence on the wall of the tube opposite the cathode, whatever may be the position of the anode. The fluorescence is even stronger if observed on a screen of zincblende placed within the tube opposite the cathode. The fluorescence is said to be caused by "cathode rays". The fact that the fluorescence spot may be deviated by allowing the rays to pass through an electric or magnetic field is of prime importance in obtaining a knowledge of the nature of these rays. The direction of deviation is such as to indicate that the rays are composed of negatively charged particles. We obtain the same rays if we use as a cathode a wire filament heated to incandescence by an electric current. Moreover, with this cathode, we still obtain the rays when the tube has been pumped down to pressures at which it would no longer pass a current with a cold cathode. This method of generating cathode rays is to be preferred, as we shall see presently, and we shall assume such an arrangement to be used in the experiments discussed below. Let the anode be a plane metal plate having a small hole in it and placed opposite the cathode. The greater part of the rays from the cathode strikes the anode. A current-measuring instrument con-

\* Naturally, on account of unavoidable errors of measurement, only small multiples can be demonstrated.

nected between the electrodes indicates the passage of a current; thus *the cathode rays transport electric charges*. Let the narrow beam which passes through the hole in the anode strike a fluorescent screen some distance away. The path of the ray between anode and screen may be influenced by applying electric or magnetic fields. In analogy with electrolysis, we assume the cathode rays to consist of particles of mass  $m$  (still undetermined) and charge  $e$ . We note at once one important difference between the two phenomena: if we reverse the potential applied to the tube no current flows. The particles effecting the transport of electricity in this instance are all of one charge, negative.

We can calculate at once the velocity of the particles from the conservation of energy:

$$eV = \frac{1}{2} m v_0^2, \quad . . . . . (4)$$

or

$$v_0 = \sqrt{\frac{2e}{m} V}. \quad . . . . . (4')$$

The values of  $e$  and  $m$  are still unknown; we seek to obtain the necessary relationships from experiments on electrical and magnetic deflection of the particles. The deflection by an electric field may be calculated very simply. Let the field be normal to the original direction of motion, which we take as the  $x$ -axis of a rectangular system of co-ordinates. Each particle is then subject to a constant force  $eE$  in the  $y$ -direction. The problem is thus identical with the computation of the trajectory of a particle thrown horizontally in a gravitational field. Starting with

$$\left. \begin{aligned} m\ddot{x} &= 0 \\ m\ddot{y} &= eE \end{aligned} \right\}, \quad . . . . . (5)$$

we integrate twice and apply the initial conditions, obtaining

$$\left. \begin{aligned} x &= v_0 t \\ y &= \frac{eEt^2}{2m} \end{aligned} \right\}. \quad . . . . . (6)$$

Eliminating the time, the trajectory is found to be the parabola

$$y = \frac{eE}{2mv_0^2} x^2. \quad . . . . . (7)$$

The electrical deviation after passing a distance  $a$  through the field of strength  $E$  is then

$$h_{el} = \frac{eEa^2}{2mv_0^2} = \frac{Ea^2}{4V}, \quad . . . . . (8)$$

i.e. by inserting the value of  $v_0$  from (4'),  $e/m$  drops out of the equation,



Hence the electrical deflection alone gives only the potential  $V$ , which is directly measurable in our case.

We seek to obtain a further equation from the magnetic deflection. In order to simplify the arrangement we apply the field along the  $z$ -axis in a direction normal to the initial direction of motion of the particles. On p. 309 we calculated the force exerted by a magnetic field on an element of a conductor carrying a current in a vacuum to be

$$\mathbf{F} = \frac{1}{c} [i\mathbf{H}] d\tau.$$

But a volume element containing charge of density  $\rho$  and moving with velocity  $\mathbf{v}$  is equivalent to a portion of a convection current of density  $\rho\mathbf{v}$  (cf. p. 418). The charge  $e$  is equivalent to  $\rho d\tau$ , so the force on the element of current becomes

$$\mathbf{F} = \frac{\rho}{c} [\mathbf{v}\mathbf{H}] d\tau = \frac{e}{c} [\mathbf{v}\mathbf{H}]. \quad . . . . . (9)$$

It follows that the force is always normal to the vector  $\mathbf{v}$  along the orbit, hence the magnetic field can do no work: it can change the direction of  $\mathbf{v}$  but not its magnitude, so that the kinetic energy is not changed. Further, since  $\mathbf{F}$  is also normal to  $\mathbf{H}$ , no acceleration occurs along the  $z$ -axis. Thus the  $z$ -component of the velocity remains zero if—as we assume—it was zero when the particle entered the field. On the other hand, the field imparts a  $y$ -component of velocity to the particle. We have

$$\mathbf{v} = \frac{dx}{dt} \mathbf{i} + \frac{dy}{dt} \mathbf{j} \quad \text{and} \quad \mathbf{H} = H\mathbf{k}, \quad . . . (10)$$

so that the component forms of the equation of motion are

$$\left. \begin{aligned} m \frac{d^2x}{dt^2} &= \frac{e}{c} H \frac{dy}{dt} \\ m \frac{d^2y}{dt^2} &= -\frac{e}{c} H \frac{dx}{dt} \end{aligned} \right\} . . . . . (11)$$

Integrating once and applying the initial condition that the velocity has the direction of the  $x$ -axis and the magnitude  $v_0$  when  $x = y = 0$ , we obtain

$$\left. \begin{aligned} \frac{dx}{dt} &= \frac{e}{mc} Hy + v_0 \\ \frac{dy}{dt} &= -\frac{e}{mc} Hx \end{aligned} \right\} . . . . . (12)$$

Squaring and adding these two equations and noticing that the magnitude of the velocity is constant, we obtain the following equation of the trajectory:

$$x^2 + y^2 + \frac{2mcv_0}{eH} y = 0. \quad . . . . . (13)$$

This is the equation of a circle touching the  $x$ -axis at the origin and of radius

$$\rho = \frac{mcv_0}{eH}. \quad \cdot \cdot \cdot \cdot \cdot \cdot (14)$$

The paths of charged particles are bent into circular arcs in a magnetic field normal to the initial direction of motion. These circles lie in a plane normal to the field and their radii are directly proportional to the initial velocity and inversely proportional to the strength of the field. If the magnetic field is of limited extent such that only the segment  $b$  of the  $x$ -axis is in the field, and if  $H$  is not too great,  $y^2$  may be neglected as compared with  $y$  in the equation of the path, and we find the magnetic deflection to be given by

$$h_{\text{mag}} \approx \frac{eHb^2}{2mcv_0} = \frac{Hb^2}{2c\sqrt{2V}} \sqrt{\frac{e}{m}}, \quad \cdot \cdot \cdot \cdot (15)$$

where  $v_0$  is replaced by its value from equation (4') (p. 427). This approximation is equivalent to neglecting the change in the components of  $v$  caused by the field itself in the right members of (11), i.e. to setting  $dy/dt = 0$  and  $dx/dt = v_0$ . It is not possible to determine  $v_0$ ,  $e$  and  $m$  separately, since we have only two equations, but equation (4) (p. 427) and equation (15) above represent two equations for  $v_0$  and  $e/m$ . The numerical value of  $e/m$  thus found is

$$\frac{e}{m} = 5.273 \times 10^{17} \text{ e.s.u./gm.} = 1.759 \times 10^{11} \text{ coulomb/kg.,}$$

i.e. 1840 times as great as the ratio of charge to mass for the lightest electrolytic ion, the  $H$  ion. Since we always find the same value of  $e/m$  regardless of whether we use a heated cathode or a cold one, and regardless of the nature of the residual gas in the tube, it is natural to interpret the result by identifying  $e$  with the magnitude of the electrolytic element of charge and thus conclude that the particles in the cathode rays are 1840 times smaller in mass than the hydrogen atom. This interpretation is the only logical one, for if we were to assume the particles to have a mass equal to that of the atom of residual gas, it would be difficult to explain the fact that  $e/m$  is independent of the kind of gas. We have thus disclosed the existence of a new fundamental substance, the atom of negative electricity, which we call the *electron*. Using the value of  $e$  obtained from the Faraday equivalent charge and Avogadro's constant, the mass of the electron is found to be

$$m_0 = 9.11 \times 10^{-28} \text{ gm.}$$

The mass of the hydrogen atom is  $1.67 \times 10^{-24}$  gm.

It is a fact of greatest importance that more accurate measurements show the ratio  $e/m$  to be not exactly constant, but to decrease as the velocity of the particles increases. Kaufmann was the first to succeed in demonstrating this variability. He applied transverse electric and magnetic fields to a beam of electrons moving in the

same direction but with various velocities in such way that the electric and magnetic deviations were at right angles to each other. Take a co-ordinate system in the plane of the observing screen, the origin being at the point where the undeviated beam meets the screen, with the  $x$ - and  $y$ -axes in the directions of electric and magnetic deviation respectively. Then, according to equation (8) (p. 427) and equation (15) (p. 429),

$$x = \frac{eEa^2}{2mv_0^2}, \quad y = \frac{eHb^2}{2mcv_0}.$$

If we look upon  $v_0$  as a continuously variable parameter, these are the parametric equations of a curve traced upon the screen by the electronic rays of various initial velocities. Elimination of  $v_0$  yields

$$x = \frac{2mc^2Ea^2}{eH^2b^4} y^2,$$

a parabola passing through the origin and having as axis the  $x$ -axis. Even if the experimental curve does not extend quite to the origin, it is possible to determine the tangent at the origin (which corresponds to infinite velocity) by reversing the electric field. The result of the experiment was, however, that the tangent formed a small but unmistakable angle with the  $y$ -axis. In addition to  $e/m$ , the equation of the curve contains only constants of the apparatus. The discrepancy, is, then, interpretable only by assuming that  $e/m$  varies, and in such way that it decreases as  $v_0$  increases. Now the relativity theory provides for an increase of inertial mass with increasing velocity according to the formula (cf. p. 254)

$$m = \frac{m_0}{\sqrt{1 - \frac{v^2}{c^2}}},$$

while there is no reason whatsoever for assuming any variation of  $e$ . As a matter of fact, Kaufmann's results are in good accord with this formula. Abraham deduced a formula for the variability of mass, based upon the concept of a fixed ether, which gives slightly different departures. More exact experiments, however, gave results decidedly in favour of the relativity formula (see p. 478). The relationship between potential drop  $V$  and velocity must also be modified from the relativistic point of view. The correct energy equation replacing equation (4) (p. 427), is

$$m_0c^2 \left( \frac{1}{\sqrt{1 - \frac{v^2}{c^2}}} - 1 \right) = eV. \quad . \quad . \quad . \quad (16)$$



Finally, all  $e/m$  determinations must be reduced to zero velocity; this has already been done for the numerical values given above.

*Ex. 111.* In deflection experiments the screen is generally not exactly at the end of the field, but there is a field-free space between it and the edge of the field. Show that one may treat the deflections observed on the screen as if the rays were bent at the middle of the field by the angle which the tangent to their path at the end of the field makes with the original direction of motion.

*Ex. 112. Theory of the "Magnetic Lens":* A slightly divergent beam of electrons emanates from a point source, the vertex angle of this cone of rays being  $2\alpha$ . Assume all the electrons to have the same speed  $v$ . Show that a magnetic field parallel to the axis of the cone will re-unite the rays approximately at a point a certain distance away from the source.

### 3. Survey of the Possible Methods of Generating Carriers of Electricity in Gases.

#### *Preliminary remarks concerning the structure of matter and the interaction of light and matter*

At the present time we are certain beyond doubt that the atom is a complicated configuration of electric charges. The entire mass of the atom, with the exception of the negligible mass of the outer electrons, is concentrated in the positively charged "nucleus" whose diameter— $10^{-12}$  cm.—is vanishingly small compared with that of the whole atom (kinetic theory radius  $10^{-8}$  cm.). This nucleus is surrounded by a shell of electrons which fill a space of about the size of the kinetic theory volume. In an electrically neutral atom the number of electrons outside the nucleus is equal to the resultant number of positive elementary charges in the nucleus, and this number is identical with the *atomic number*, i.e. the number giving the place of the element in the periodic table. An atom which has lost one or more of its outer electrons as a result of some external influence represents a positive ion, while an atom with an extra electron is a negative ion. A system having several nuclei is called a molecule. A molecule, as well as an atom, may become either a positive or a negative ion. It is chiefly the halogens, which come just before the noble gases in the periodic system, which tend to form negative ions. These elements form the singly-charged anions  $F^-$ ,  $Cl^-$ ,  $Br^-$ ,  $I^-$ .

In a metal in the solid or liquid state, some of the electrons are able to move freely between the atoms. The state of this "electron gas" may be compared with that of an actual gas dissolved in a solid or a liquid.

To remove a single electron from an atom or from the surface of a metal requires a definite amount of energy which is characteristic of the particular substance. This *work of escape* is generally specified by giving the potential difference through which an electron must fall in order to acquire this amount of energy. The magnitude of this

*ionization potential* for free atoms varies from 3.88 volts for cæsium to 24.48 volts for helium. Once an electron has attained the critical speed it is able to dislodge an electron from the electron shell of an atom with which it collides. This happens relatively frequently as soon as the electrons have reached the critical speed, while the ionizing collisions of fast electrons are less frequent. The impact of a positive ion can also result in ionization of an atom, but the impinging particle must have a considerably greater speed, for according to the laws of impact, there is a considerable transfer of kinetic energy of translation when two approximately equal masses collide. But even for speeds theoretically sufficient for ionization, the number of effective collisions is minimal. In contrast with electronic impact, the number does not increase until the velocities become greater. Instead of a positive ion, a neutral atom may impinge, and again ionization will occur under certain conditions. Since neutral atoms usually possess velocity as a result of thermal energy of agitation, this process, which is of importance at temperatures of 3000° C. and beyond, is called *thermal ionization*. There are thermodynamic formulæ which make possible the calculation of the degree of ionization of a gas at a given temperature and pressure.

Finally, radiation of short wave-length (<3000 Å.) also causes ionization. *In all energy exchanges between atoms and radiation, light acts as if it were composed of discrete quanta of energy, of amount proportional to the frequency of the light, which are transferred as units.* The factor of proportionality is denoted by  $h$  and is Planck's celebrated *quantum of action*, whose numerical value is

$$h = 6.62 \times 10^{-27} \text{ erg sec.}$$

In order that there be ionization by light—*photo-ionization*—the incident "light quantum"  $h\nu$  must be at least equal to the work of ionization. Calling the latter  $eV_i$ , we have

$$h\nu \geq eV_i. \quad . \quad . \quad . \quad . \quad . \quad . \quad (17)$$

If we characterize the light by its frequency instead of by its wave-length, insertion of the numerical values yields the easily-remembered relationship between wave-length and equivalent electron energy

$$\frac{\lambda}{(\text{In Å.U.})} \times \frac{V_i}{(\text{in volts})} = 12380.$$

Atoms struck by electrons whose speed is not sufficient to cause ionization may nevertheless be made to emit radiation provided that the electrons have fallen through a potential exceeding a value called the *excitation potential*  $V_e$ . Then equation (17) holds in reverse, i.e.

$$h\nu \leq eV_e. \quad . \quad . \quad . \quad . \quad . \quad . \quad (17')$$



All the agents here described can also dislodge electrons from the surface of a metal. Thus the impact of electrons or ions on a metallic surface releases "secondary electrons"; light incident on a metal releases "photo-electrons"; heat causes "thermal emission of electrons".\* The last-mentioned phenomenon is of particular interest. Just as an absorbed gas escapes from the absorbing body when heat is applied, electrons may be "evaporated" from a metal by heating it to around  $2000^{\circ}\text{C}$ . The emission of electrons takes place at a lower temperature ( $600^{\circ}\text{C}$ .) if the metal is coated with a thin layer of alkaline earth oxides.

For completeness we must mention the possibility of the formation of ions in the course of chemical reactions. An example is the ionization of air by slow oxidation of phosphorus or other substances.

#### 4. The Separately-sustained Electrical Discharge. Spark Discharge.

A self-sustaining discharge is one in which the passage of the current itself continues to produce the necessary carriers of the charge, while a discharge in which these agents are produced entirely or in part by an external source of ionization, so that the discharge ceases when the source is turned off, may be referred to as separately sustained.

The simplest example of a separately-sustained discharge is pure electronic conduction in a high vacuum. The electrons are liberated at the cathode either by radiation (photoelectric cell) or by heat, and move to the anode under the influence of the applied field. The maximum value of the current is attained when all the electrons liberated from the cathode each second reach the anode. This current, which cannot be increased by raising the potential difference between anode and cathode, is called the *saturation current*. On the other hand, for small negative values of the potential difference between the cold electrode and the hot one, a very few high-speed electrons will succeed in crossing the tube against the field. The voltage-current curve must therefore have the general appearance shown in fig. 1. The rigorous derivation for the linear case (extended parallel plate electrodes) was given by Gans. The theory is somewhat complicated, since it is not correct to assume a linear variation of potential from cathode to anode (which would, of course, be valid if the cathode were cold) on account

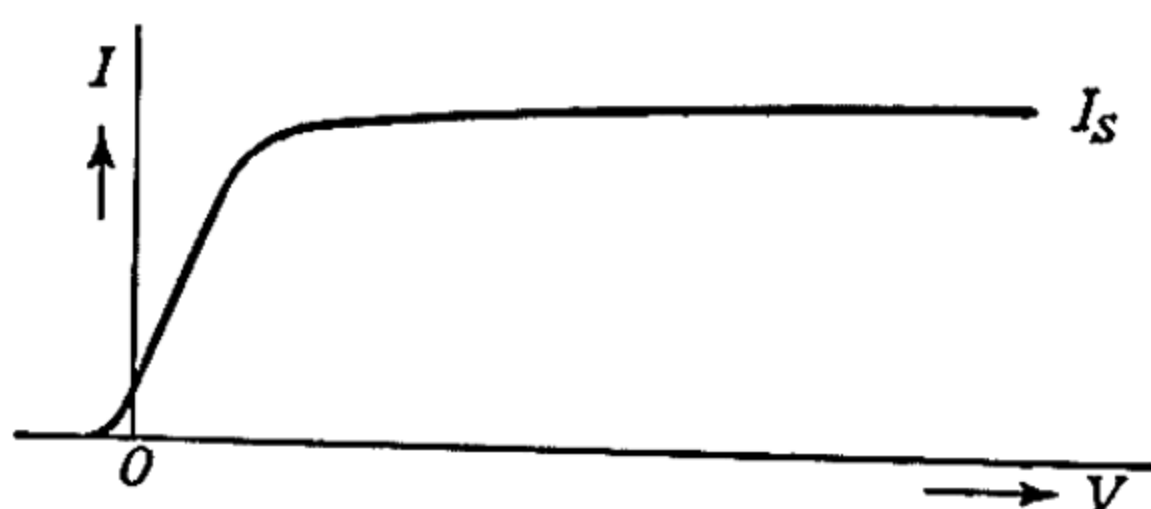


Fig. 1

\* The work of escape from a metal is less than that required to free an electron from the atom.



of the distortion of the field by the cloud of electrons surrounding the hot cathode. The fact that no appreciable current passes when the potential difference is negative permits the use of the hot-cathode tube as an electrical valve (rectifier). If a wire grating or grid through which most of the electrons pass is interposed between anode and cathode, the electron current arriving at the anode depends very strongly on any difference of potential applied between the grid and the cathode. As a result, even a small alteration in grid potential causes the anode current to change considerably. This property—the ability of the tube to act as a regulator and amplifier—makes the three-electrode tube one of the most important devices in wireless telegraphy and telephony.

We now consider the case in which the gas between the electrodes is artificially ionized, say by X-rays, while nothing unusual is assumed to occur at the electrodes. In this case, too, we may expect a saturation current, viz. when all the ions formed in the gas in unit time reach the electrodes without previously recombining with oppositely-charged ions. If  $q$  ion pairs are formed in each cubic centimetre per second, and if the electrodes are  $l$  centimetres apart, the density of the saturation current is  $i_s = qle$ . This seems paradoxical at first sight: the greater the separation of the electrodes, the greater the current. It is to be remembered, however, that this result refers to the *highest attainable* current, and that the potential must be raised when the electrode separation is increased in order to maintain the saturation current.

We find experimentally that a new current increase occurs when the potential is raised further. This is due to the formation of new secondary ion pairs by the primary ions and electrons. Instead of treating this somewhat more complicated case, we shall investigate the simpler instance in which the primary electrons are formed at the cathode only, rather than in the gas. By means of light or some other agent, let  $n_0$  electrons be liberated from each square centimetre of the cathode every second. Then the saturation current will have the density  $i_s = n_0e$ . Let each electron form  $a$  secondary ion pairs per centimetre of its path. Assume the velocity of the positive ions to be insufficient for ionization. If, in addition to the  $n_0$  primary electrons,  $n_1(x)$  secondary electrons pass through a cross-section at  $x$  each second—a total of  $n(x)$  electrons—then between  $x$  and  $x + dx$

$$dn = nadx \quad . \quad . \quad . \quad . \quad . \quad . \quad (18)$$

additional electrons are released each second. Integrating between  $x = 0$  and  $x = l$ , we obtain the total number of electrons arriving at the anode

$$n = n_0e^{al} \quad \text{i.e.} \quad i = en_0e^{al}. \quad . \quad . \quad . \quad . \quad (19)$$

This same value of the current density is equal to the total charge on the positive secondary ions arriving at the cathode plus that of the departing negative primary electrons. The form of equation (19) shows that the photoelectric current may be increased enormously by filling the cell with gas.

We observe no sharp limiting current because there is a second possibility of ionization when the potential is increased, viz. ionization by collision of positive ions with atoms. Assume each positive ion to form  $\beta$  ion pairs per centimetre of path. Let  $n_0$  primary electrons be released from each square centimetre of the cathode per second by some external agent, and let the total number of electrons arriving at the anode be  $n_1$ . Since no positive ions are liberated at the anode,  $n_1$  is also the measure of the total current, which, at the cathode, consists of the electron current plus the ion current. Considering again a thin layer between  $x$  and  $x + dx$  (fig. 2), there are  $n$  pairs of ions formed to the left of  $x$  per second and—neglecting infinitesimals— $(n_1 - n)$  pairs formed on the right side. For a saturation current at equilibrium we must also have  $n$  electrons crossing the left face of the layer from the left, per second, and  $(n_1 - n)$  positive ions crossing the right surface from the right. Hence the number of electrons released on the path  $dx$  per second amounts to

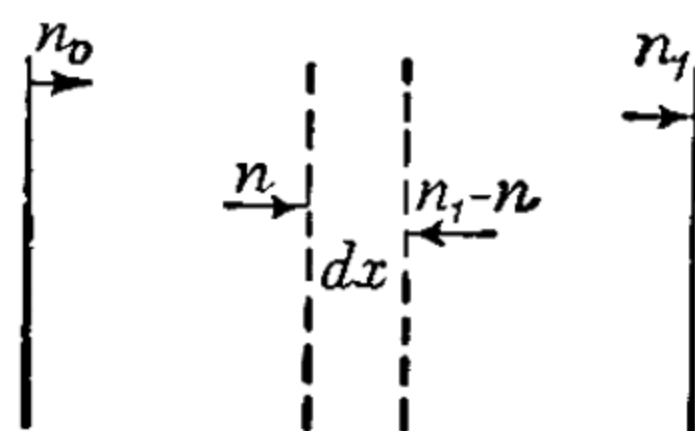


Fig. 2

$$dn = [n\alpha + (n_1 - n)\beta]dx = [n(\alpha - \beta) + n_1\beta]dx, \quad (20)$$

whence, by integration,

$$n(\alpha - \beta) + n_1\beta = Ce^{(\alpha - \beta)x}. \quad (21)$$

We must have  $n = n_0$  when  $x = 0$ , so that

$$C = n_0(\alpha - \beta) + n_1\beta. \quad (22)$$

The single constant of integration is thus determined. The boundary condition for  $x = l$  leads to an equation for  $n_1$ :

$$n_1(\alpha - \beta) + n_1\beta = [n_0(\alpha - \beta) + n_1\beta]e^{(\alpha - \beta)l}. \quad (23)$$

Solved for  $n_1$ , this is

$$n_1 = \frac{n_0(\alpha - \beta)e^{(\alpha - \beta)l}}{\alpha - \beta e^{(\alpha - \beta)l}}. \quad (23')$$

It is interesting to notice that the denominator becomes zero if

$$\alpha = \beta e^{(\alpha - \beta)l} \quad \text{or} \quad l = \frac{\log \alpha / \beta}{\alpha - \beta}. \quad (24)$$

In this case even a very small value of  $n_0$  is accompanied by a very



large (mathematically infinite) value of  $n_1$ . Then the function of the primary electrons is only to initiate the process known as *electrical breakdown* or *spark discharge*.

The ionization coefficients  $\alpha$  and  $\beta$  will be proportional to the number of collisions on unit path and also to the kinetic energy acquired between two collisions. Since the number of collisions is proportional to the pressure, while the energy acquired on the free path is proportional to the ratio of the field strength  $E$  to the pressure  $p$ ,\* we can write

$$\alpha \propto p\phi\left(\frac{E}{p}\right), \quad \beta \propto p\psi\left(\frac{E}{p}\right). \quad \cdot \cdot \cdot \cdot (25)$$

Remembering that  $E = V/l$  for plane electrodes, and designating by  $q$  a factor of proportionality which does not contain either  $V$  or  $p$  or  $l$ , equation (24) yields

$$lp = q \frac{\log \frac{\phi(V/lp)}{\psi(V/lp)}}{\phi(V/lp) - \psi(V/lp)}. \quad \cdot \cdot \cdot \cdot (26)$$

In place of the ionization coefficients we now have functions of  $V$ ,  $l$  and  $p$  in the relationship for the spark discharge. Considering this an equation for the spark potential  $V$ , we see that for plane electrodes *the spark potential is proportional to the product of electrode separation and gas pressure*, i.e. to the amount of gas between the electrodes. This law, which was discovered empirically by Paschen, holds also for spherical electrodes if the radii of the spheres are increased in the same ratio as their distance apart.

Paschen's law is experimentally well substantiated, but the starting mechanism assumed here is not the only one that leads to this result. In fact, according to present knowledge, ionization by positive ions plays only a minor role, while indirect release of electrons from the metal of the cathode is much more important. The latter process comes about in this way: An electron not only causes ionization but—even before it has enough energy to do this—it excites the emission of light from atoms with which it collides. This radiation releases electrons from the cathode photoelectrically, since the work of escape from the metal is less than that from free atoms. The condition for initiation of the discharge is that each primary electron shall liberate at least one electron from the cathode. Let  $\gamma$  be the average number of electrons released indirectly from the cathode by one electron. Then, since the number of secondary electrons liberated in a distance  $l$  by a single primary electron is, by the above,  $e^{\alpha l} - 1$ , we must require that

$$\gamma(e^{\alpha l} - 1) \geq 1. \quad \cdot \cdot \cdot \cdot (27)$$

\* For the energy is proportional to the product of field strength by free path, and the latter is inversely proportional to the pressure.



Since the same thing is true for the excitation of atoms to produce radiation as was said above for the coefficients  $\alpha$  and  $\beta$ , we see that  $\gamma$ , too, is independent of  $V/lp$  and Paschen's law holds for this starting mechanism also.

This assumed process seems to be valid up to pressures of about 200 mm. of mercury. Cloud chamber photographs show that at higher pressures the discharge is initiated by "canals" which propagate themselves with the aid of irregularities in the field produced by the space charge due to positive ions (see next section).

### 5. Self-maintaining Discharge: Glow and Arc Discharges.

In the foregoing calculations the distortion of the field by space charge was neglected in examining the initiation of discharge. Whether this is justified seems doubtful at the present time (see preceding

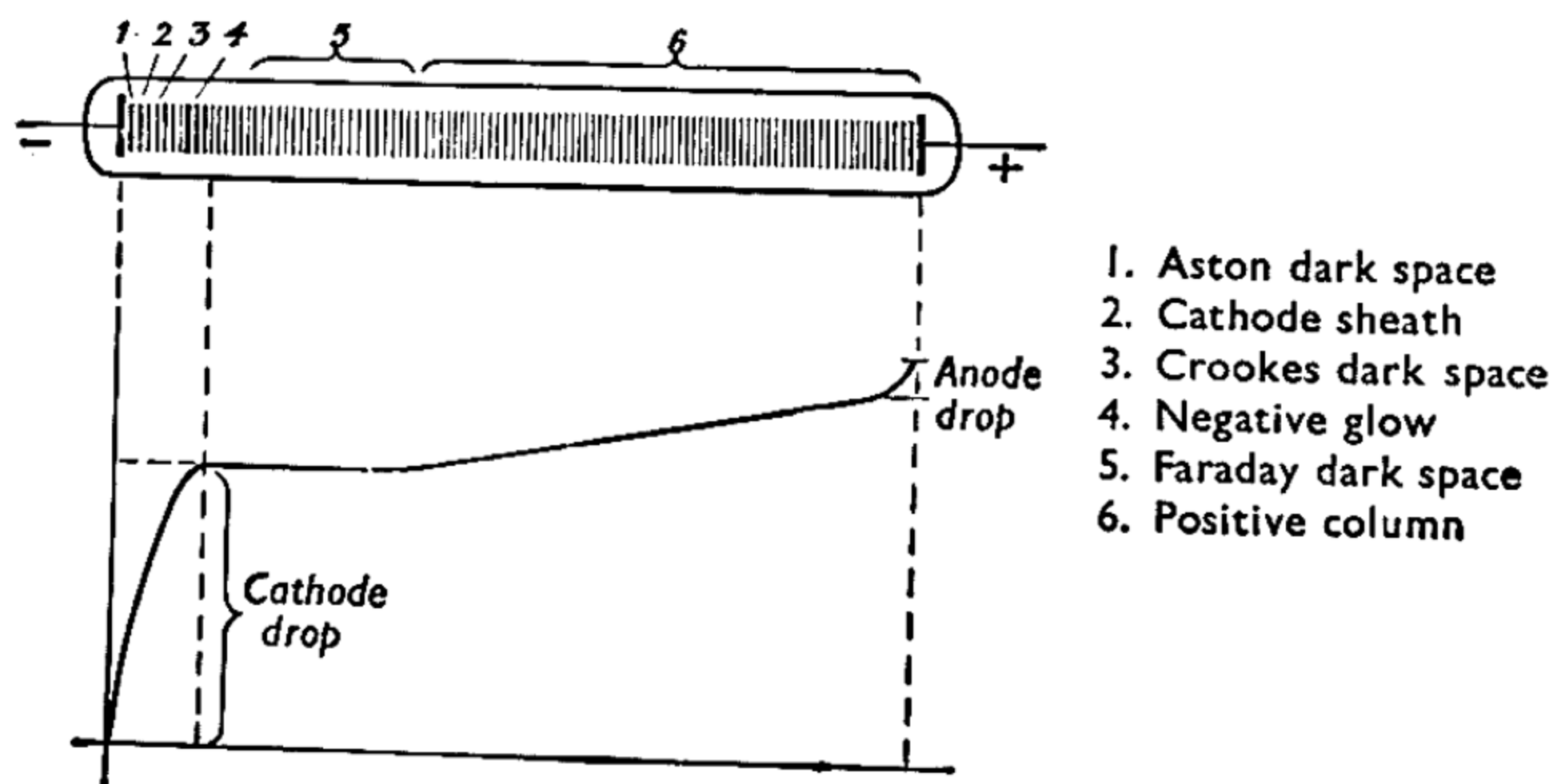


Fig. 3

section). At any rate, the steady forms of discharge, the main types of which are the glow discharge and the arc, are characterized by the fact that, on account of the large space charge, the field changes considerably after the discharge begins. In the present section we shall examine the properties of a stationary discharge without entering into a consideration of the very difficult question of what circumstances cause a disruptive discharge to pass over into a steady one.

Fig. 3 represents the main features of the potential variation in a "glow discharge", which occurs when low currents traverse a tube filled with gas at low pressure. The various segments of the potential curve manifest themselves in part by different regions of luminosity. The nomenclature is given in fig. 3. The explanation, however, is very complicated, and will not be attempted here. The rapid change of potential at the cathode is called the *cathode drop*. Quantitatively, this

is defined as the difference in potential between the cathode and the point of minimum field strength, i.e. least potential gradient. This potential difference is *independent* of the current strength and of the gas pressure and depends only upon the nature of the gas and of the metal of the cathode, provided the current remains below a certain value. This limiting condition may be recognized visually by the fact that the cathode is just covered by the negative glow at that stage. The order of magnitude of the normal cathode drop is 200 volts. The cathode drop is the vital feature of a glow discharge; it corresponds to approximately the lowest potential with which a discharge can be maintained. What, then, is the source of the great lack of symmetry between cathode and anode? It is based upon the difference between the electron and the positive ion; this difference is responsible for quite different conditions for positive and negative space charges. On account of the relationship

$$eV = eEl_f = \frac{1}{2}mv^2. \quad . \quad . \quad . \quad . \quad . \quad (28)$$

( $l_f$  = mean free path), an electron attains a velocity  $\sqrt{1840A}$  times as large as a positive ion of atomic weight  $A$  in falling through the same potential difference. Moreover, on account of the vanishingly small size of the electron, its free path  $l_f$  is greater than that of the positive ion. Thus the electron engages in fewer collisions and so can accumulate more energy. According to the kinetic theory of gases, the mean free path of an atom of radius  $r_0$  projected into a cluster of stationary atoms of radius  $r_1$ , of which the number per cubic centimetre is  $n_1$ , is given by

$$l_f = \frac{1}{\pi(r_0 + r_1)^2 n_1}. \quad . \quad . \quad . \quad . \quad . \quad (29)$$

If these atoms are also in motion with velocities approaching that of the impinging atom, the factor  $1/\sqrt{2}$  must be supplied in the right member. The velocity attained by the ions is of the order of magnitude of the gas-kinetic velocity, so that the second assumption is valid for them; the electrons, on the other hand, move so much faster that we can consider the gas atoms relatively at rest. Hence for the electron ( $r_0 = 0$ ) the mean free path is

$$l_e = \frac{1}{\pi r_1^2 n_1}, \quad . \quad . \quad . \quad . \quad . \quad (30)$$

while for an ion,

$$l_i = \frac{1}{4\pi r_1^2 n_1 \sqrt{2}}. \quad . \quad . \quad . \quad . \quad . \quad (31)$$

In a given field the velocity of the electrons is to that of the ions as

$$\sqrt{1840A \cdot 4\sqrt{2}} : 1, \quad \text{or as} \quad 102\sqrt{A} : 1.$$

The region of the discharge which is free of space charge is characterized by linear potential rise, for Poisson's equation (p. 268) demands that  $d^2V/dx^2$  vanish where there is no space charge. Hence, at such places,

$$n_+ = n_-$$

$$i = i_+ + i_- = en_+v_+ + en_-v_-,$$

so that

$$\frac{i_-}{i_+} = \frac{v_-}{v_+} = 102\sqrt{A}. \quad . \quad . \quad . \quad . \quad . \quad (32)$$

In the region of no space charge, i.e. essentially in the positive column, the portion of the current carried by the electrons is at least one hundred times that carried by the positive ions.

If we now assume for simplicity that the current at the cathode is carried only by positive ions, that at the anode by electrons alone, and that the field is the same on both sides, then—since  $i = Nev$ —the ratio of the space charges  $Ne$  before cathode and anode is at least  $\sqrt{102A} : 1$ . Therefore, on account of Poisson's equation

$$\frac{d^2V}{dx^2} = -4\pi\rho = -4\pi Ne, \quad . \quad . \quad . \quad . \quad . \quad (33)$$

the potential curve must turn sharply downward at the cathode and slightly upward at the anode. Naturally, these considerations are only qualitative; in reality, part of the current at the cathode is carried by electrons dislodged from the metal by the impact of ions which have attained considerable speed in traversing the cathode fall. Precisely this liberation of electrons is necessary for the maintenance of the discharge.

In the absence of an external resistance for limiting the current, the glow discharge can go over into the arc form, in which the current is several hundred times greater. The arc is distinguished from the glow by the fact that the heating of the cathode causes it to release a much greater electron current that would be furnished by positive ion impact alone. Thus the arc type of discharge presumes the use of a hot cathode.\*

In many arcs the temperature of the gas is so high that thermal ionization is of considerable importance in addition to ionization by electron impact. It follows from these remarks that the number of charge-bearing particles increases considerably as the current is made greater, since both the thermal emission of the cathode and the

\* There are certain forms of electric arc which do not employ thermal emission. In these the electron emission takes place under the influence of very strong electric fields.



thermal ionization of the gas increase exponentially with the temperature. This causes the arc to have a so-called "*falling characteristic*"; the resulting potential difference  $V$  across the arc and the current  $I$  are related by an equation of the form

$$V = a + \frac{b}{I}, \quad . . . . . (34)$$

which is the exact opposite of the ohmic relationship  $V = IR$ .

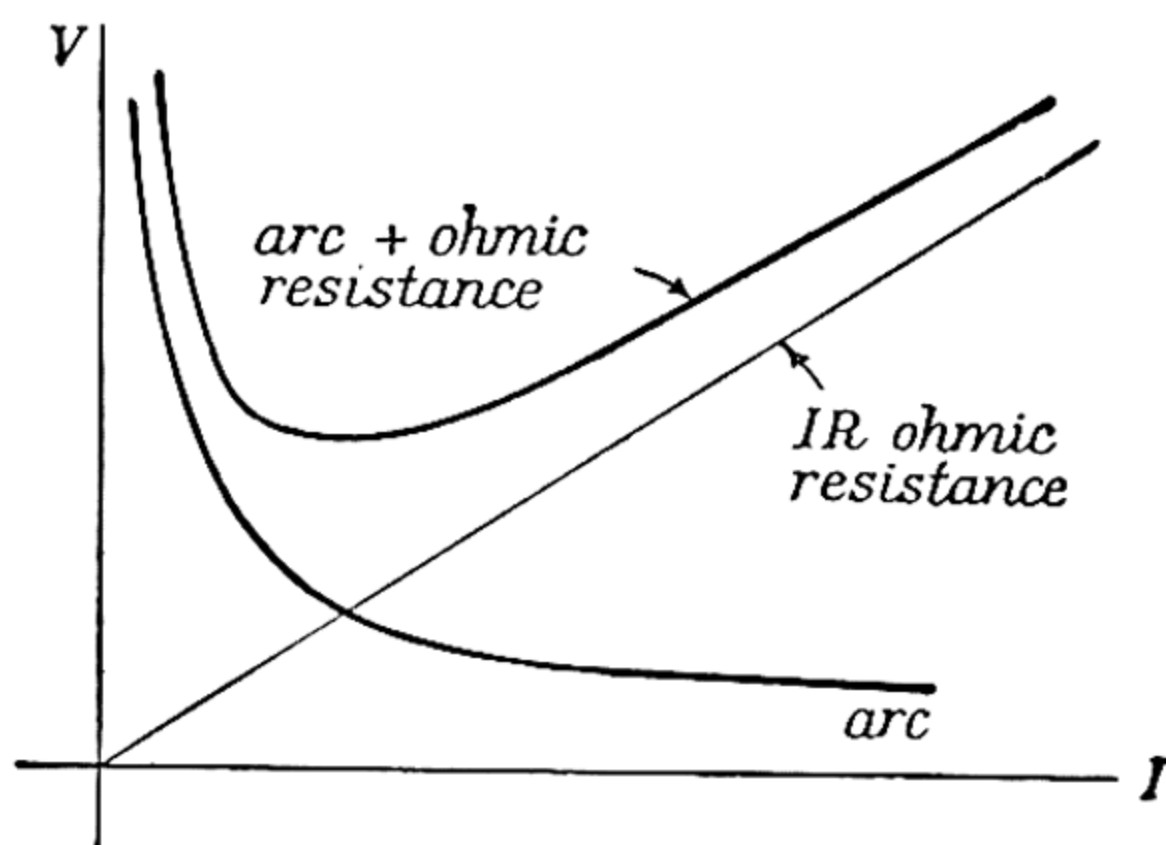


Fig. 4

A conductor having a falling characteristic  $V = f(I)$  must not be connected directly to a source of potential, as may be seen by considering fig. 4. If, through some cause, the current should increase, the E.M.F. opposing the battery potential diminishes; the current then increases, causing the potential to drop further, resulting in an additional increase in current. On the other hand, for a conductor with a rising characteristic (e.g. an ohmic resistance), the required potential rises along with the current, so that the current returns to the equilibrium value, since the battery is incapable of giving the required E.M.F. when the current increases. One must therefore make certain, by inserting a conductor having a rising characteristic, that the characteristic of the entire system is an increasing one.

## 6. The Origin of Cathode, Canal, and Positive Rays. The Mass Spectrograph.

Having obtained a broad view of the phenomena of the glow discharge, we are able to understand the origin of the various rays which are found in a discharge tube under favourable conditions. It must be remembered that there is a strong field near the cathode, and that the positive ions attain a considerable velocity in this field. The cathode rays, recognizable at low pressures by their fluorescent effects, are a part of the electrons released from the cathode by positive ion impact, viz. those which have traversed the entire potential fall without appreciable energy loss by collision. But in order to maintain the discharge, it is necessary that a certain number of the liberated electrons produce new electrons and ions by impact in the gas.

The positive ions which have attained high speeds in the cathode drop can be made visible in the form of rays by providing the cathode with small holes and making the space behind the cathode field-free. Part of the ions then pass through the holes in the form of "canal rays", where they gradually lose their speed by impacts with the randomly-moving particles of the gas. The principal difference between canal rays and cathode rays is that the positive ions constituting the former are self-luminous, while the electrons are capable only of causing other substances to radiate. The Doppler Effect (p. 235) may be perceived in the light of the canal rays, as first recognized by J. Stark. As in the case of cathode rays, the velocity  $v$  and specific charge  $e/M$  may be determined by electric and magnetic deflection. The results show—as may be expected—that the speeds are not alike, since the distances traversed by the various particles vary considerably.

Rays of positive ions can be generated at the anode also if conditions are so arranged that a greater anode fall results, and if the anode is made of a material capable of emitting positive ions, e.g. a metallic salt. As mentioned above, the smallness of the anode fall is due to the high speed of the light electrons. If now the inertia of the negative charges is artificially increased by attaching them to atoms, the resulting field at the anode must be similar to that before the cathode. This loading of the electrons may be promoted by introducing electronegative atoms into the gas, i.e. such atoms as the halogens, which have a tendency to form negative ions. This is done in practice by using a halogen salt as anode; the cations of the salt pass off as *positive rays* while the halogen evaporates as a neutral atom.

The determination of  $e/M$  for canal rays and for positive rays is of particular importance, for it makes possible the determination of the atomic weight of individual ions. It is assumed, of course, that the charge on the particle is known; this must consist of an integral number of elementary charges, and there is usually no doubt as to the exact number. We see from equation (8) (p. 427) and equation (15) (p. 429) that the electric deflection is the same for all particles which have traversed the same potential drop, while the magnetic deflection depends upon  $e/M$  as well as upon the potential difference  $V$ . This circumstance may be used to overcome the troublesome lack of homogeneity of the canal ray particles which, of course, have fallen through various potential differences. By means of an electric field, the particles are first sorted according to  $V$ ; then, by means of a slit  $S_2$  (fig. 5), a narrow region containing only ions which have traversed the same potential drop is isolated; finally, by means of a magnetic field, the particles are sorted according to mass. This is the scheme of F. W. Aston's "mass spectrograph". The sharpness of separation depends upon the homogeneity of the rays, which, in turn, depends upon the

narrowness of the slit; this is limited by the fact that the intensity of the beam must not be made too low by undue reduction of the slit width. Aston's idea was to focus the rays in such manner that rays within a given range of potential difference would be united at the photographic plate which serves to record them. Integration of equa-

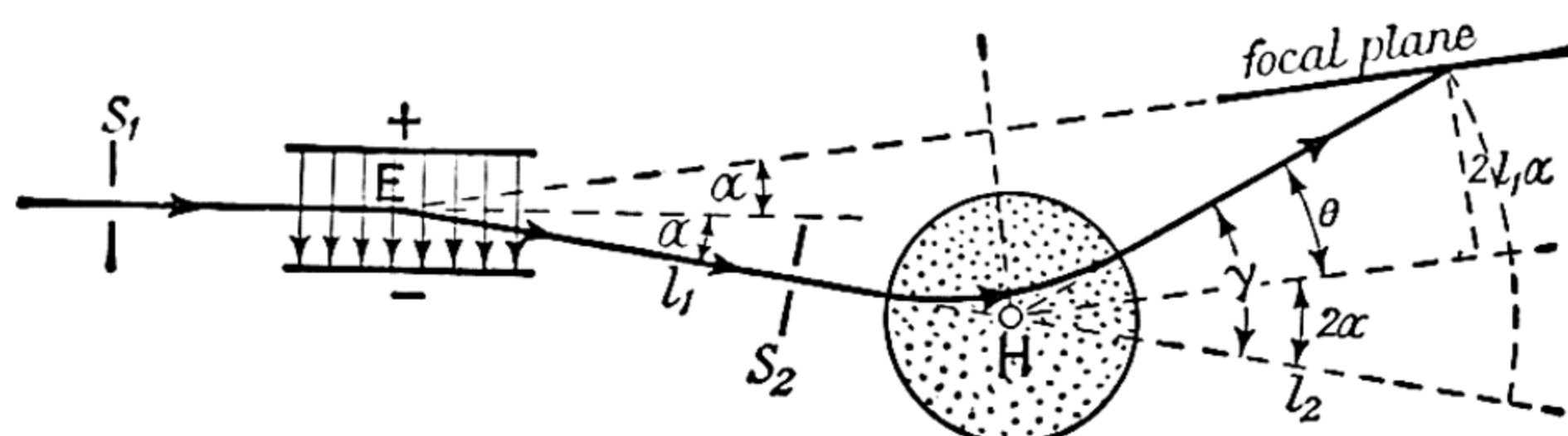


Fig. 5

tion (5) (p. 427) gives the angular deviation in an electric field of length  $a$ , provided the angle is small:

$$\alpha = \frac{dy}{dx} = \frac{eEa}{Mv_0^2} = \frac{Ea}{2V}, \quad \dots \dots (35)$$

and the magnetic angular deviation in a field of length  $b$  is, by (15), p. 429,

$$\beta = \frac{dy}{dx} = \sqrt{\frac{e}{M}} \cdot \frac{Hb}{c\sqrt{2V}}. \quad \dots \dots (36)$$

Call the distance between the mid-points of the electric and magnetic fields  $l_1$  and let the distance travelled by the rays from the middle of the magnetic field to the detector be  $l_2$ . According to *Ex. 111* (p. 431), we may consider the rays to come in a straight path from the centre of the field. The entire deflection, measured in a plane normal to the original direction of motion, is then given by

$$h = (l_1 + l_2)\alpha + l_2\beta. \quad \dots \dots (37)$$

If it happens that the potential drop traversed by the rays is not strictly constant, but varies by an amount  $dV$ , then the angles of deflection lie in an interval

$$d\alpha = -\alpha \frac{dV}{V} \quad \text{and} \quad d\beta = -\beta \frac{dV}{2V}, \quad \dots \dots (38)$$

as may be seen by logarithmic differentiation of equations (35) and (36). This means that rays within this interval which have fallen through different potential differences but have the same  $e/M$  form a pencil of cross-section

$$dh = \left[ (l_1 + l_2)\alpha + \frac{l_2\beta}{2} \right] \frac{dV}{V}. \quad \dots \dots (39)$$



This cross-section is reduced to zero by making the expression in brackets vanish. This can happen only when  $\alpha$  and  $\beta$  have opposite signs, i.e. when the magnetic deflection is opposite in direction to the electric. Set  $\beta = -\gamma$ . Since  $l_1$  is fixed, and since  $\alpha$  also has a fixed value for a given potential, the only remaining variable is the distance  $l_2$  from the mid-point of the magnetic field to the point where the rays are received. The bracketed expression set equal to zero gives the equation of the focal curve in polar co-ordinates  $l_2, \gamma$ —i.e. the curve on which the pencils corresponding to various values of  $e/M$  have zero cross-section. This equation is

$$l_2 = \frac{2\alpha l_1}{\gamma - 2\alpha} \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad (40)$$

The polar axis is the direction of the magnetically undeviated ray (see fig. 5). If we choose a new axis making an angle  $2\alpha$  with this one, the angle measured from the new axis is  $\theta = \gamma - 2\alpha$ , and the equation of the focal curve takes the simple form

$$l_2 \theta = l_2 \sin \theta = 2l_1 \alpha. \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad (41)$$

This is, in fact, a straight line parallel to the new polar axis and distant from it  $2\alpha l_1$ . This line is the intersection of the photographic plate with the plane of the figure when the plate is properly placed to receive the rays of a given  $e/M$  in sharp focus. The magnetic field has a dual function: it separates the rays corresponding to various masses and unites (focusses) the rays corresponding to particles having the same mass but originating at different points in the cathode drop.

This "classical" example of a mass spectrograph, since surpassed by other designs, is one of the first instances of the use of *electron optics*. The term "focussing" has been current ever since. Electron optics is based on the formal correspondence between light paths in various refracting media and particle orbits in fields of force (cf. p. 787 *et seq.*). Following this analogy through, individual fields of force may be computed as "lenses" (as in *Ex. 112*, p. 431), which are the basic elements of electron-optical instruments such as the well-known electron microscope.

## CHAPTER XXIV

### THE FUNDAMENTAL IDEAS OF THE THEORY OF METALLIC CONDUCTION

#### 1. Electrons as Carriers of Current in Metals.

There are only a few solids which, like  $\text{CuI}$ , exhibit good electrolytic conductivity. The current in such substances is carried by material ions, as may be seen by the deposits formed at the electrodes. The conductivity of that class of substances known as metals is many orders of magnitude greater—in fact, an essential characteristic of a metal is high electric conductivity unaccompanied by any transport of matter. It is therefore likely that the current is carried by electrons which pass through the meshes of the atomic or ionic lattice. From the validity of Ohm's Law for metals we conclude, as in the case of electrolytes, that the reaction between the electrons and the lattice particles may be looked upon as a frictional force  $r\mathbf{v}$  proportional to the velocity, so that the relationship

$$-e\mathbf{E} = r\mathbf{v} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

is again valid.\* The direct proof that the current is carried by electrons in metals was given by Tolman in his determination of the value of  $e/m$  for the conduction electrons. Since an electron has inertial mass, this ratio must be determined from experiments in which the inertia of the electrons manifests itself. One might imagine, for example, an arrangement in which a metallic conductor is freely suspended. If the conductor is at rest while a current is flowing, and if the current is suddenly interrupted, the electrons continue to possess a certain amount of momentum which is destroyed by friction within the body. According to the principle of the Conservation of Momentum, the body as a whole must then experience a displacement. The opposite effect shows more promise of verification, for practical reasons: mechanical acceleration of the conductor should cause relative motion of the electrons, and therefore a current. Since we have very sensitive instruments for measuring feeble currents, this method is easier to

\* The negative sign arises from the negative charge of the electron.

apply; Tolman actually determined  $e/m$  in this way (1914-23). We examine the case of an infinite straight cylindrical conductor which experiences a constant acceleration  $\alpha$  in the direction of its length. According to p. 231, each electron is subject to an inertial force  $-m\alpha$ , and under the influence of this force and friction it attains a velocity

$$v = -\frac{m\alpha}{r}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

If there are  $N$  (negative) electrons per cubic centimetre, the density of the resulting current will be

$$i = +\frac{Nem}{r}\alpha. \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

Now, on account of the relationship

$$i = -Ne v,$$

or

$$i = \frac{Ne^2 E}{r}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

since we have always

$$i = \sigma E, \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

we find

$$\sigma = \frac{Ne^2}{r}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

Hence the current which results from the acceleration is given by the equation

$$i = \frac{\sigma m}{e} \alpha. \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

Assume now that in place of a straight conductor we have one of circular form that we accelerate by oscillating it around an axis normal to the plane of the circle. While it is true that the acceleration is no longer constant, (7) may be applied at any instant, understanding  $\alpha$  to be the instantaneous acceleration. The direction of  $\alpha$ , and hence of  $i$ , is understood to be along the tangent. We then have circular currents, and the oscillation results in a magnetic field which varies with time. Its induction effects on a secondary coil may be computed by the results of p. 312, exactly as for a normally produced alternating current. Since (7) contains  $e/m$  as the only unknown, Tolman was able to calculate its numerical value. Within the limits of error, it agreed with that for free electrons.

The next question is: "What is the state of the conduction electrons in metals?" The simplest and most fruitful assumption is that the electrons move about through the framework of atoms, like atoms of





be taken into account. If there are  $N$  free electrons per cubic centimetre, the magnitude of the current density is

$$i = -Ne\bar{v}_x = \frac{e^2 N l E}{2m v_w} = \sigma E. \quad . \quad . \quad . \quad (11)$$

Since none of the factors composing  $\sigma$  depends upon the field  $E$ , we have here the proportionality between current density and field strength, which is equivalent to Ohm's Law.

With regard to the temperature dependence of  $N$ ,  $l$  and  $v_w$ , we may say the following: According to our ideas concerning the structure of metals,  $N$  must in any case be temperature-dependent. Further, considering the electrons to form a monatomic gas, we have, according to the statistical law of equipartition (cf. p. 596), on the average, at a temperature  $T$

$$\frac{1}{2} m v_w^2 = \frac{3}{2} k T, \quad . \quad . \quad . \quad . \quad (12)$$

where  $k$  is the Boltzmann constant, or "gas constant per molecule",  $R/L$ . Thus  $v_w$  increases with  $\sqrt{T}$ . If we make use of the experimental fact that the conductivity of all pure metals is inversely proportional to the absolute temperature, we can conclude that  $l$  decreases with  $1/\sqrt{T}$ , so that the product  $l v_w$  is independent of the temperature. Qualitatively, we can see at once that the temperature affects each of these quantities oppositely: an increase in temperature causes the lattice elements to vibrate with greater amplitude. This increases the number of collisions between electrons and atoms, causing a reduction of the mean free path. On the other hand, the mean thermal velocity increases. However, according to the newer theory (cf. next section),  $v_w$  must be essentially independent of temperature and  $l$  must decrease as  $1/T$ .

### 3. Conduction of Heat in Metals; the Law of Wiedemann and Franz.

The fact that the thermal conductivity of metals is many orders of magnitude greater than that of solid dielectrics suggests that this property, too, is determined by the free electrons. If this is so, we can compute the thermal conductivity of metals on the same basis of an electron gas and bring it into relation with the electrical conductivity. For the sake of simplicity, we again ignore the statistical distribution of velocities and of free paths and use the single magnitude  $v$ , corresponding to the quantities denoted by  $v_w$  above, and the single value  $l$ . On the other hand, we are not justified in assuming that the *components* of  $v$  are equal for all electrons. Assume a temperature gradient in the direction of the  $z$ -axis. Consider the electrons passing in both directions through a cross-section at a height  $z$ . We assume that the

energy of these electrons corresponds to that of the layer in which they experienced their last collision. If we denote the energy of an electron at height  $z_0$  by  $u_0$ , and call the temperature  $T_0$ , then

$$u(z) = u_0 + \left(\frac{du}{dz}\right)_0 z = \frac{3}{2} kT_0 + \frac{3}{2} k \frac{dT}{dz} z. \quad . \quad . \quad (13)$$

The electrons passing through each square centimetre of the section  $z_0$  every second transport a total amount of energy

$$S = \Sigma n_{v_z} v_z \left[ u_0 + \left(\frac{du}{dz}\right)_0 z \right], \quad . \quad . \quad . \quad (14)$$

where  $n_{v_z}$  is the number of electrons per cubic centimetre which have the  $z$ -component of velocity  $v_z$ . The summation is to be extended over all  $v_z$  which occur, and over all possible layers  $z$ . These, however, are not independent. If we have uniform velocity  $v$  and if we replace the distance between the plane  $z = 0$  and the place of the last collision by the mean free path, we have

$$\cos(kv) = \cos \theta = \frac{v_z}{v} = \frac{z}{l}.* \quad . \quad . \quad . \quad (15)$$

Again, the expression  $\Sigma n_{v_z} v_z$  vanishes, otherwise there would be an electric current  $-e \Sigma n_{v_z} v_z$  flowing through the section  $z = z_0$ . The thermal energy transported each second is

$$S = \Sigma n_{v_z} v \frac{z^2}{l} \frac{du}{dz} = \Sigma \frac{3}{2} \frac{kn_{v_z} v z^2}{l} \frac{dT}{dz} = \frac{3}{2} kNvl \frac{dT}{dz} \overline{\cos^2 \theta}. \quad (16)$$

Now the spatial average of  $\cos^2 \theta$  is  $1/3$ ; thus the energy flux is given by

$$S = \frac{1}{2} kNlv \frac{dT}{dz}. \quad . \quad . \quad . \quad . \quad (17)$$

The thermal conductivity is defined by

$$S = \sigma_w \frac{dT}{dz},$$

and so

$$\sigma_w = \frac{1}{2} kNlv. \quad . \quad . \quad . \quad . \quad (18)$$

Here again we do not know the manner in which the separate factors

\* It seems at first remarkable that we may put the mean free path  $l$  for the distance between the planes  $z = z$  and  $z = z_0$ , since all the electrons do not experience their next collision in the plane  $z = z_0$ . This will be discussed in more detail on p. 570.



depend upon the temperature. However, if we form the ratio to the electrical conductivity, a simple expression results:

$$\frac{\sigma_w}{\sigma_{el}} = \frac{1}{2} \frac{kNlv \cdot 6kT}{e^2Nlv} = 3 \left( \frac{k}{e} \right)^2 T = 3 \left( \frac{R}{F} \right)^2 T. \quad (19)$$

This expresses the law of Wiedemann-Franz, which states that *the ratio of thermal to electrical conductivity is proportional to the absolute temperature*. This law is well substantiated experimentally, and the factor of proportionality agrees, within certain limits, with the value  $3(R/F)^2$ .

#### 4. Objections to the Theory Developed Above. The Electron Theory of Pauli and Sommerfeld.

Although the concept of an electron gas accounts for many phenomena of metallic conduction, the fact that some of its consequences lead to gross contradictions of experience must not be overlooked. Primarily, there is the question of the contribution of the electrons to the energy content, and hence to the specific heat, of the metal. According to the Law of Equipartition of Energy (p. 596), there corresponds to every degree of freedom which contributes to the energy of a system an amount of energy  $kT/2$  per molecule. Assuming as above that the number of free electrons is of the same order of magnitude as the number of atoms, the energy content of the metal must be increased by  $(3/2)RT$  per mol, corresponding to the three components of the translatory motion of the electrons. Then the specific heat per mol,

$$c_v = \left( \frac{\partial u}{\partial T} \right)_v$$

does not have the value  $6R/2$  found for higher temperatures (Law of Dulong and Petit), but amounts to  $9R/2$ , which is completely at variance with experience. This was first cleared up by recognizing that the electron gas has an enormous number of particles per unit volume—one electron per atom, or as many individuals per unit volume as there are atoms of the *solid* metal. Thus it is not legitimate to treat an electron gas by ordinary statistics. The theory developed by Pauli and Sommerfeld will be discussed later (Chap. XXXVIII).

## CHAPTER XXV

### ELECTRON THEORY OF THE DIELECTRIC CONSTANT, INDEX OF REFRACTION AND MAGNETIC PERMEABILITY

#### 1. The Origin of Electrical and Magnetic Polarization.

While the dielectric constant and the permeability (or the electric displacement vector **D** and the magnetic induction vector **B**) are the quantities used in the field theory, they have no immediate physical interpretation in the atomistic theory of electrical phenomena. It is the polarization, i.e. the resultant dipole moment per cubic centimetre, whose interpretation can be given by the atomistic theory. The polarization **P**—in the magnetic case, the magnetization **M**—depends upon **D** and **E** or upon **B** and **H** according to the relationships given on pp. 280, 298:

$$\mathbf{D} = \mathbf{E} + 4\pi\mathbf{P}, \quad \mathbf{B} = \mathbf{H} + 4\pi\mathbf{M}.$$

The polarization vector is parallel to the field vector for isotropic bodies; for anisotropic substances (crystals not belonging to the regular system) **P** is a linear vector function of **E**. We restrict our present considerations to isotropic substances, for which the connexion between **P** and **E** or between **M** and **H** is given by a simple scalar magnitude—the susceptibility  $\kappa$ .

A resultant dipole moment can originate in two ways in either the electrical or the magnetic case. Firstly, the molecules of the substance may themselves possess a moment. In this instance the axes of the dipoles are randomly oriented, so that there is no resultant moment. Application of an external field causes the particles to assume the direction of minimum energy, i.e. to set themselves parallel to the field. This adjustment, however, is opposed by the thermal agitation, so that the alignment can become perfect only for an infinitely strong field or when the temperature is reduced to absolute zero. It is evident that this kind of susceptibility depends strongly upon the temperature. In the magnetic case the term “paramagnetic” has been applied for a long time; in the electrical case the differentiation is less obvious (see below), and so this distinction has come to be applied only recently. The corresponding term “paraelectric” or “parelectric” is therefore not yet in universal use.



The other possibility is that the dipole moment of an individual molecule is induced by the field itself. This process is easy to comprehend in the electrical case. The electronic shells are not rigidly attached to the core, so that an electric field causes a relative displacement of the two. On account of the fact that the charge on the electron is negative, the shells are moved in a direction *opposite* to the field, but since the dipole moment is defined as the limit of the product of charge by separation, multiplied by the unit vector drawn in the direction from the negative toward the positive charge, the moment itself has the direction of  $\mathbf{E}$ . Hence the resulting "dielectric polarization" is in the direction of  $\mathbf{E}$ . In the electrical case the dielectric and piezoelectric polarizations have the same sign. However, *since the dielectric polarization does not involve molecular orientation, it does not depend on the temperature.* Dielectric polarization occurs also for piezoelectric substances, since a relative displacement of positive and negative charges can occur here also. The distinction between the two forms will become clearer in the following paragraph.

The origin of "diamagnetism" is somewhat more complicated. As shown on p. 664, an applied magnetic field imparts to the electrons an additional rotation about the direction of the field. This so-called Larmor precession takes place with angular velocity  $\omega = eH/2mc$ . This motion of the electrons gives rise to a magnetic moment which, as exact calculation shows, is in a direction opposite to the field. Thus the diamagnetic and paramagnetic susceptibilities are opposite in direction, and so the distinction between the two forms in the magnetic instance is the well-known one. The quantum theory requires that the paramagnetic moment of an atom be a multiple of a certain magnitude, the magneton; this is so large that for paramagnetic substances the diamagnetic susceptibility which is always present is relatively insignificant.

## 2. Theory of Dielectric Polarization, Optical Index of Refraction and Dispersion.

Assume that the substance under consideration contains  $N$  atoms or molecules per cubic centimetre, these particles themselves having no dipole moment. Suppose further that the distances between particles are so great that their mutual influences may be neglected. In a field  $\mathbf{E}$  every particle assumes a moment, which we may take as proportional to the field, for we are dealing with small displacements of the electronic systems, and may take the restoring forces to be proportional to the displacements. These forces hold the system in equilibrium in the field; hence the elongation, and thus the dipole moment, is proportional to the field strength. We then have

$$\mathbf{p} = a\mathbf{E}, \quad \dots \dots \dots (1)$$



and the polarization becomes

$$\mathbf{P} = \kappa \mathbf{E} = N\alpha \mathbf{E}. \quad . . . . . (2)$$

The dielectric constant is then given by

$$K = n^2 = 1 + 4\pi N\alpha, \quad . . . . . (3)$$

where the Maxwell relationship  $K = n^2$  is assumed (p. 331). Inasmuch as we shall limit our present considerations to matter of very low density, we may replace  $(n^2 - 1)$  by  $2(n - 1)$  and obtain an approximate formula for the refractive index:

$$n - 1 \approx 2\pi N\alpha. \quad . . . . . (3')$$

Since we assume a quasi-elastic restoring force, the application of an alternating field will cause displacements of the electron shells which depend upon the frequency. Assuming further that the polarization of each atom is caused by the displacement of an individual electron from its position of equilibrium, the equation of motion of the electron is that of a forced harmonic vibration:

$$m \frac{d^2 s}{dt^2} + ks = -eE_0 e^{i\omega t}.* \quad . . . . . (4)$$

The stationary solution is, according to p. 97,

$$s = -\frac{eE_0 e^{i\omega t}}{k - m\omega^2}, \quad . . . . . (5)$$

or introducing the natural frequency  $\omega_0 = \sqrt{k/m}$ ,

$$s = -\frac{eE_0 e^{i\omega t}}{m(\omega_0^2 - \omega^2)}. \quad . . . . . (6)$$

Hence the dipole moment becomes

$$\mathcal{P} = \mathcal{P}_0 e^{i\omega t} = \frac{(e^2/m)E_0 e^{i\omega t}}{\omega_0^2 - \omega^2}, \quad . . . . . (7)$$

and the polarizability is given by

$$\alpha = \frac{e^2/m}{\omega_0^2 - \omega^2}. \quad . . . . . (8)$$

If the number of electrons is  $f$  instead of one, the value is  $f$  times as great and the refractive index is given by

$$n - 1 = \frac{fNe^2/m}{2\pi(\nu_0^2 - \nu^2)} \quad . . . . . (9)$$

where  $\omega$  has been replaced by  $2\pi\nu$ .

\* The effect of the magnetic vector of the light wave, given by equation (9) (p. 428) may be neglected, since the resulting electron velocities are small with respect to  $c$ , while  $\mathbf{E}$  and  $\mathbf{H}$  are of the same order of magnitude in our system of measurement.

Even this simple dispersion formula contains all the essential properties of the optical index of refraction. The characteristic frequency  $\nu_0$  is to be interpreted as the *absorption frequency* of the atom, since for this applied frequency the response of the atom is so great that all energy is taken from the primary ray and scattered in every direction. Now the absorption lines of free atoms lie almost exclusively in the ultra-violet region of the spectrum. Hence if we trace the course of the index in the visible region,  $\nu < \nu_0$ , and so the refractive index is always greater than unity. As we pass to shorter wave-lengths the frequency increases and we approach the characteristic frequency; the difference  $\nu_0^2 - \nu^2$  becomes smaller and smaller and the index increases toward short waves. This is the usual behaviour for glass and for liquids, and is called *normal* dispersion. Beyond the absorption band the index becomes less than unity according to equation (9), and so the rays in that region are refracted less than the longer ones. This reversal of the usual prismatic sequence of colours was first observed by Christiansen, using fuchsine, and is called *anomalous* dispersion, although from the point of view of the theory there is no anomaly whatsoever.

The dispersion formula may be improved in several respects. Firstly, the fact that an atom or molecule may possess not one, but perhaps a large number of absorption lines must be taken into consideration. Each one contributes an oscillator whose strength is proportional to the intensity, so that we write

$$n - 1 = \frac{Ne^2/m}{2\pi} \sum_i \frac{f_i}{(\nu_{0i}^2 - \nu^2)} \cdot \cdot \cdot \cdot \quad (10)$$

Moreover, the oscillations of the quasi-elastically bound electrons are not undamped; the radiation resulting from the oscillation (see p. 341) involves a certain damping which prevents the index from actually becoming infinite at the absorption lines, as in equation (10). If the damping is taken into account, the index becomes complex and the substance behaves like a metal near the regions of absorption. We shall not treat this case in more detail here.

It is more important to take into consideration the mutual interactions of the dipoles, which cannot be neglected when we consider denser substances, such as liquids or solids. In such cases there is a microscopic field  $\mathbf{F}$  at the place occupied by any selected electron. This field is not the same as the macroscopic field  $\mathbf{E}$  defined, for example, by the difference of potential across a condenser. But in any case the susceptibility is defined as the ratio of the polarization to the *macroscopic* field strength. The difference between  $\mathbf{E}$  and  $\mathbf{F}$  is that *all* the charges contribute to the macroscopic field, while in calculating the field  $\mathbf{E}$ , acting upon the dipole under consideration, the latter is not to be included (cf. footnote, p. 266). In order to calculate

the field of neighbouring dipoles, which is added to the applied field, we circumscribe a small sphere about the point occupied by the atom under consideration and examine first the contribution of the dipoles lying outside the sphere—i.e. we imagine all material removed from within the sphere. There is a surface charge upon the sphere, and the additional field has the value  $+4\pi\mathbf{P}/3$ , according to p. 283, where we had the same surface charge, but with opposite sign (see also *Ex. 75*, p. 284). According to a calculation made by Lorentz, the contribution of the particles within the sphere vanishes either when there is complete absence of order or when the atom under observation is a point of a regular crystal lattice, i.e. for all isotropic bodies. Then the dipole moment of the individual atom or molecule is

$$\mathbf{p} = \alpha \left( \mathbf{E} + \frac{4\pi}{3} \mathbf{P} \right), \quad . . . . . (11)$$

and we obtain the equation

$$\mathbf{P} = N\alpha \left( \mathbf{E} + \frac{4\pi}{3} \mathbf{P} \right) \quad . . . . . (12)$$

for the polarization. Solving this equation for  $\mathbf{P}$  yields

$$\mathbf{P} = \kappa \mathbf{E} = \frac{K-1}{4\pi} \mathbf{E} = \frac{N\alpha \mathbf{E}}{1 - \frac{4\pi}{3} N\alpha}. \quad . . . . . (13)$$

Solving for  $N\alpha$  shows that the simple relationship between  $N$ ,  $\alpha$  and  $K$  (or  $n^2$ ) given by equation (3) (p. 452) is to be replaced by

$$\frac{4\pi}{3} N\alpha = \frac{K-1}{K+2} = \frac{n^2-1}{n^2+2}. \quad . . . . . (14)$$

Hence for denser substances the dispersion formula is

$$\frac{n^2-1}{n^2+2} = \frac{Ne^2/m}{3\pi} \sum_i \frac{f_i}{\nu_{0i}^2 - \nu^2}. \quad . . . . . (15)$$

As may be verified immediately, this formula becomes identical with equation (3) for  $n \approx 1$ . It is customary to deal with one mol of the substance, rather than with 1 c.c. Let the density be  $\rho$  and the molecular weight  $M$ . Then 1 gm. contains  $L/M$  molecules, and 1 c.c. contains

$$N = \frac{L\rho}{M}. \quad . . . . . (16)$$

Using this expression in equation (15), we obtain

$$\frac{n^2-1}{n^2+2} \cdot \frac{M}{\rho} = \frac{Le^2/m}{3\pi} \sum_i \frac{f_i}{\nu_{0i}^2 - \nu^2}. \quad . . . . . (17)$$



The quantity on the left, which has the dimensions of a volume, is called the *molecular refraction*.

Extrapolation of equation (15) to infinitely long waves, i.e. to  $\nu = 0$ , gives the following equation for the static value of the dielectric constant:

$$\frac{K - 1}{K + 2} = \frac{Ne^2/m}{3\pi} \sum_i \frac{f_i}{\nu_{0i}^2} \cdot \cdot \cdot \cdot \cdot \quad (15')$$

For many substances this does yield the true value, but for others—e.g. water—there are large discrepancies. The reason is that this method gives only the dielectric part of the polarization. As will be seen in the next section, the paelectric portion vanishes for optical frequencies, on account of the fact that the charges are not able to follow the field at such frequencies. Thus substances for which (15') gives the correct value of  $K$  are free of dipoles.

*Ex. 113.* The *refractivity* of a substance of mass  $m$  and density  $\rho$  is, in analogy with the molecular refraction given by the expression

$$R = \frac{n^2 - 1}{n^2 + 2} \cdot \frac{m}{\rho}.$$

Show that the refractivity of a mixture is the sum of the refractivities of the components.

*Ex. 114.* Formerly, the molecular refraction was often taken to be the actual volume occupied by the molecules. As a matter of fact, one obtains numerical values which agree well with the kinetic theory values. Show that this is so, for example, for water ( $n = 1.33$ ). Show also that the numerical values of polarizability are those of metallic spheres of the gas-kinetic size (of. p. 284).

*Ex. 115.* Calculate phase velocity and group velocity of electro-magnetic waves in a medium containing  $N$  free electrons per unit volume (Heaviside Layer of the upper atmosphere).

### 3. Paelectric Susceptibility.

As explained in § 1 (p. 450), the application of a field to a substance whose molecules possess an electric moment generates polarization by lining up these dipoles. This phenomenon occurs only if the molecules are free to move—i.e. principally in gases and in liquids—and if the external field reverses slowly enough to permit the system to come to equilibrium between reversals. The paelectric contribution, which is responsible for the high value of the dielectric constant of water, vanishes at frequencies corresponding to wave-lengths of a few centimetres. In order to find the extent to which the dipoles are aligned in terms of the temperature and field strength we make use of the statistical theorem (p. 588) that the number of particles of a system which are in a state of energy  $u$  is proportional to

$$e^{-u/kT},$$

where  $k$  is the Boltzmann constant. Now the potential energy of a dipole  $\mathbf{p}$  in a field  $\mathbf{E}$  is given by

$$u = -\mathbf{p}\mathbf{E} = -pE \cos \theta. \quad . \quad . \quad . \quad (18)$$

If we denote the direction of the dipole axis by a point on the unit sphere, the number of these points which lie in the zone  $\theta$  to  $\theta + d\theta$  ( $\theta =$  co-latitude) is proportional both to the area of the zone and to

$$e^{-u/kT}.$$

We may therefore set

$$dN = C e^{pE \cos \theta / kT} \sin \theta d\theta, \quad . \quad . \quad . \quad (19)$$

the factor  $2\pi$  being contained in  $C$ . The value of this constant is determined by integrating to obtain the total number  $N$  of particles per unit volume:

$$N = C \int_0^\pi e^{pE \cos \theta / kT} \sin \theta d\theta. \quad . \quad . \quad . \quad (20)$$

Make the substitution

$$\frac{pE \cos \theta}{kT} = w;$$

then

$$N = C \frac{kT}{pE} \int_{-pE/kT}^{+pE/kT} e^w dw = \frac{2CkT}{pE} \sinh \frac{pE}{kT}. \quad . \quad . \quad (21)$$

Solving for  $C$  and using this value in equation (19), we have

$$dN = \frac{NpE}{2kT \sinh \frac{pE}{kT}} e^{pE \cos \theta / kT} \sin \theta d\theta. \quad . \quad . \quad . \quad (22)$$

If the axis of a dipole forms an angle  $\theta$  with the field, its contribution to the resultant parallel to the field is  $p \cos \theta$ . The components normal to the field cancel out in summing over all dipoles, and we obtain the resulting dipole moment per cubic centimetre, i.e. the polarization, by integrating over all directions:

$$P = \frac{Np^2E}{2kT \sinh \frac{pE}{kT}} \int_0^\pi e^{pE \cos \theta / kT} \cos \theta \sin \theta d\theta. \quad . \quad . \quad (23)$$

Making the same substitution as above, this becomes

$$P = \frac{NpkT}{2pE \sinh \frac{pE}{kT}} \int_{-pE/kT}^{+pE/kT} e^w w dw = \frac{NpkT (we^w - e^w) \Big|_{-pE/kT}^{+pE/kT}}{2pE \sinh \frac{pE}{kT}}$$

$$\text{or} \quad P = Np \left\{ \coth \left( \frac{pE}{kT} \right) - \frac{1}{\left( \frac{pE}{kT} \right)} \right\} = NpL \left( \frac{pE}{kT} \right) \quad (24)$$

Following Langevin, to whom this theory is due, we have designated the function  $\coth x - 1/x$  by a single symbol, say  $L(x)$ . We see that  $L(x)$  approaches unity for large values of  $x$ ; a power series development shows that the function behaves like  $x/3$  for small values of  $x$ . Since  $Np$  is the maximum value (saturation value) of the polarization, attained when all the dipoles are aligned with the field, the limit must be unity for large values of the argument if the result is to have any meaning at all. For weak fields and for temperatures which are not too low—conditions which are satisfied in most experiments—the approximation  $x/3$  is sufficient and, since  $\mathbf{P}$  is always in the direction of  $\mathbf{E}$ , we obtain

$$\mathbf{P} = \frac{Np^2}{3kT} \mathbf{E}. \quad \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot (25)$$

That is, within certain limits, the paelectric susceptibility is also independent of the field, and we can apply the formulæ of § 2 if we replace  $\alpha$  by  $p^2/3kT$ . Taking account of the interactions of the dipoles, equation (14) (p. 454) then gives for the paelectric part of the dielectric constant

$$\frac{K - 1}{K + 2} = \frac{4\pi Np^2}{9kT}, \quad \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot (26)$$

and likewise the molecular polarization

$$\frac{K - 1}{K + 2} \cdot \frac{M}{\rho} = \frac{4\pi}{9} \frac{Lp^2}{kT}. \quad \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot (27)$$

In deducing the formula for the paelectric polarization we made the natural assumption that the dipoles could assume all possible orientations with respect to the field. According to the principles of the Quantum Theory, however, only a discrete set of angles of orientation occurs. Nevertheless, the strict quantum theory calculation yields only small deviations from the classical Langevin formula. Of much greater importance is the “directional quantization” in the magnetic case; this will be treated below.

On account of the fact that the adjustment cannot take place freely, solids exhibit paelectric susceptibility only in the neighbourhood of the melting-point. Nevertheless, crystals composed of ions of different signs can show a different kind of polarization which is due to relative movement of the ions as entities. This polarization also vanishes at optical frequencies.

*Ex. 116.* Given the dielectric constant  $K = 80$  for  $T = 290^\circ$  and the optical



index of refraction  $n = 1.33$ , calculate the dipole moment of the water molecule, assuming that the dielectric and piezoelectric polarizations are additive.

#### 4. Paramagnetic, Ferromagnetic and Antiferromagnetic Susceptibility.

Molecules are the only particles which can possess electric moments, since for all atoms the centre of gravity of the negative charges coincides with that of the positive charges. On the other hand, a free atom *can* have a *magnetic* moment. There is another point of difference between paramagnetism and piezoelectricity: from the fact that the paramagnetic susceptibility of salts does not change much when the salts go into solution we may conclude that *the solid state offers no hindrance to the alignment of the dipoles*.

If we now assume that all orientations are equally allowable, we need only translate the formulæ of § 3 (p. 457) into magnetic terms, with the understanding that  $p$  now represents the *magnetic* moment of the atom or molecule. Neglecting the interaction of the dipoles we have, according to equation (25) (p. 457), the following expression for the magnetization:

$$\mathbf{M} = \frac{Np^2}{3kT} \mathbf{H}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (28)$$

The susceptibility per unit volume is then given by

$$\kappa = \frac{Np^2}{3kT}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (29)$$

Division of  $\kappa$  by the density  $\rho$  gives the susceptibility per gramme; if we then multiply by the molecular weight  $G$  we obtain the important molar susceptibility

$$\chi = \frac{\kappa G}{\rho} = \frac{Lp^2}{3kT} = \frac{M^2}{3RT} = \frac{C}{T}. \quad . \quad . \quad . \quad . \quad (30)$$

Here  $M$  represents the dipole moment per mol, i.e. the moment of one mol when the dipoles are all completely aligned; and  $C$  is called the Curie constant, after P. Curie, who found the temperature relationship given by (30)\* The experimental determination of  $C$  consists in finding the temperature variation of  $\chi$ ; from this the molar magnetic moment is found by the equation †

$$M = \sqrt{3RC}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (31)$$

P. Weiss found ‡ that the molar magnetic moments of a large number of substances determined in this way were all multiples of a certain value

$$M_w = 1125.5 \text{ e.m.u.},$$

\* P. Curie, *Journ. d. Phys.*, **4**, p. 197 (1895).

† Although we now know that this evaluation is not applicable, in order to have convenient comparison between theory and experiment we quote the results in terms of "effective moments", i.e. moments as given by (30).

‡ P. Weiss, *Comptes rendus*, **152**, pp. 79, 187 (1911).

now called the Weiss Magnetron. Nevertheless, the limits of error in these experiments are quite wide. Remarkably enough, the Bohr theory of the atom also yields a magneton whose value referred to one mol is, within the limits of error, exactly five times the Weiss magneton. From the standpoint of the quantum theory—upon which Bohr's theory is based—it is not at all to be expected that we obtain a whole number of Bohr magnetons when evaluating the observed Curie constants, for it is a characteristic of the quantum theory that only a certain set of orientations of the dipole axes with respect to the field is possible. This will be discussed in further detail in the seventh part of this work (p. 665). Here we investigate only the especially important case where a moment is allowed to set itself only exactly parallel or antiparallel to the field. The distribution is governed again by Boltzmann's principle. If  $N_+$  is the number of moments aligned parallel and  $N_-$  the number antiparallel to the field, then because  $N_+ + N_- = N$ ,

$$N_+ = \frac{Ne^{\frac{pH}{kT}}}{e^{\frac{pH}{kT}} + e^{-\frac{pH}{kT}}}, \quad N_- = \frac{Ne^{-\frac{pH}{kT}}}{e^{\frac{pH}{kT}} + e^{-\frac{pH}{kT}}}, \quad \cdot \cdot \cdot \quad (22a)$$

and so 
$$M = N_+p - N_-p = Np \tanh\left(\frac{pH}{kT}\right). \quad \cdot \cdot \cdot \quad (24a)$$

In place of the function  $\coth x - (1/x)$  we have  $\tanh x$ . The trend is again typical of a saturation curve, but the initial slope is 1 instead of  $1/3$ . In what follows we need only interpret  $L(x)$  as  $\tanh x$ , and the result corresponds to the quantum theoretical special case of two allowed orientations.

The interaction of the dipoles may be taken into account by setting the effective microscope field  $\mathbf{F}$  equal to the sum of the macroscopic field and a second quantity proportional to the magnetization; in this case, however, the factor of proportionality  $4\pi/3$  derived on p. 454 does not give a correct representation of the experimental data. We therefore insert a general empirical constant  $\nu$  instead. As long as we are justified in replacing  $L(x)$  by  $x/3$  we have the expression

$$\mathbf{M} = \frac{Np^2}{3kT} (\mathbf{H} + \nu\mathbf{M}) \quad \cdot \cdot \cdot \cdot \cdot \quad (32)$$

for the magnetization. Solving for  $\mathbf{M}$ , we obtain

$$\mathbf{M} = \frac{Np^2\mathbf{H}}{3k\left(T - \frac{\nu Np^2}{3k}\right)} = \frac{Np^2\mathbf{H}}{3k(T - \Theta)}, \quad \cdot \cdot \cdot \quad (33)$$

where the quantity  $\nu Np^2/3k$ , which has the dimensions of tempera-

ture, is denoted by  $\Theta$ . The molar susceptibility then follows the Weiss law

$$\chi = \frac{M^2}{3R(T - \Theta)} \quad \cdot \cdot \cdot \cdot \cdot \quad (34)$$

If we make a graph of  $1/\chi$  as a function of the temperature, we obtain a straight line whose intersection with the axis of abscissæ gives the characteristic temperature  $\Theta$ . If a positive value is found for  $\Theta$ , there must be a real temperature at which the susceptibility attains an extremely large value. This is actually the case for a small group of substances—the ferromagnetic metals. The characteristic temperature above which these substances behave like simple paramagnetic substances is called their *Curie point*. However, for the high values of

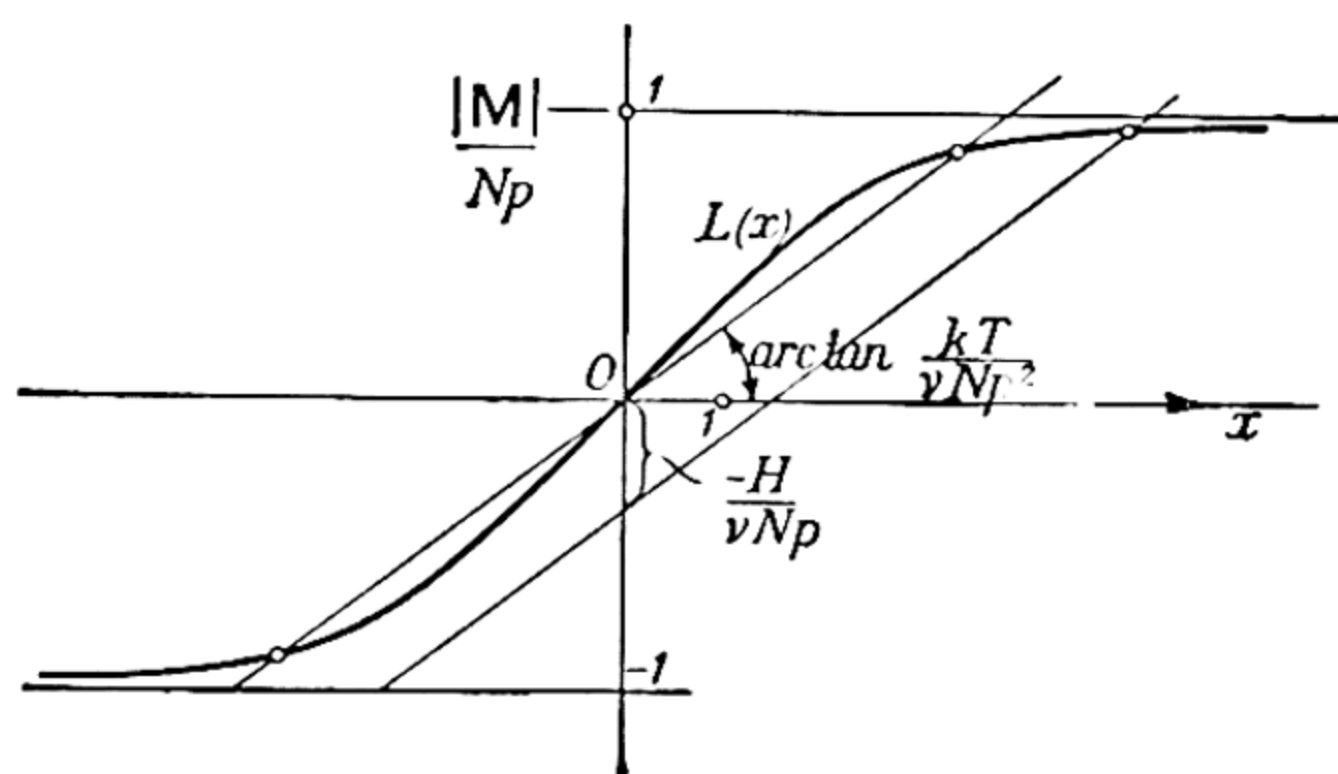


Fig. 1

ferromagnetic susceptibilities we are no longer justified in applying our approximate formula, but must turn to the Langevin function. We have the following equations for  $\mathbf{M}$  and  $\mathbf{H}$  from which  $\mathbf{F}$  is to be eliminated:

$$|\mathbf{M}| = NpL\left(\frac{pF}{kT}\right), \quad \cdot \cdot \cdot \cdot \cdot \quad (35)$$

$$|\mathbf{M}| = \frac{F - H}{v}. \quad \cdot \cdot \cdot \cdot \cdot \quad (36)$$

This may be done graphically as follows: Let the quantity  $x = pF/kT$  be the abscissa and take  $y = |\mathbf{M}|/Np$  as ordinate (fig. 1). The value of the magnetization corresponding to a given field strength  $H$  is the ordinate of the point of intersection of the curve  $y = L(x)$  with the straight line

$$y = \frac{kT}{vNp^2}x - \frac{H}{vNp}.$$

The unit along the axis of  $y$  is the saturation value  $Np$  of the mag-



netization. It is remarkable that under certain circumstances we obtain, for  $H = 0$ , two points of intersection besides  $O$ . These are symmetric about  $O$  and correspond to a high value of the magnetization. This happens when the slope of the line is less than that of the Langevin curve, i.e. less than  $1/3$ . This again gives exactly the Curie point for  $T$ . With ferromagnetic substances, a spontaneous magnetization is possible *below* the Curie point even in the absence of a macroscopic field  $H$ . Thus we have traced the phenomenon of *remanence* to the internal field. The high value of the constant  $\nu$ , however, remains a puzzle; instead of the value  $4\pi/3$  it may be of the order of magnitude of 10,000 for ferromagnetic substances.

Quantum mechanics traces  $\nu$  back to "exchange energies" (p. 712). For some substances, however, this may lead to negative values of  $\nu$ . In this instance, neighbouring spins set themselves in opposite directions as a result of the forces between them. Such *antiferromagnetic* materials (e.g. many compounds of manganese) are paramagnetic at higher temperatures. With decreasing temperature, the paramagnetism decreases in the usual way until the Curie point (here sometimes called the Néel point) is reached, at which time the alignment sets in. This is accompanied by a sudden decrease in the paramagnetism.

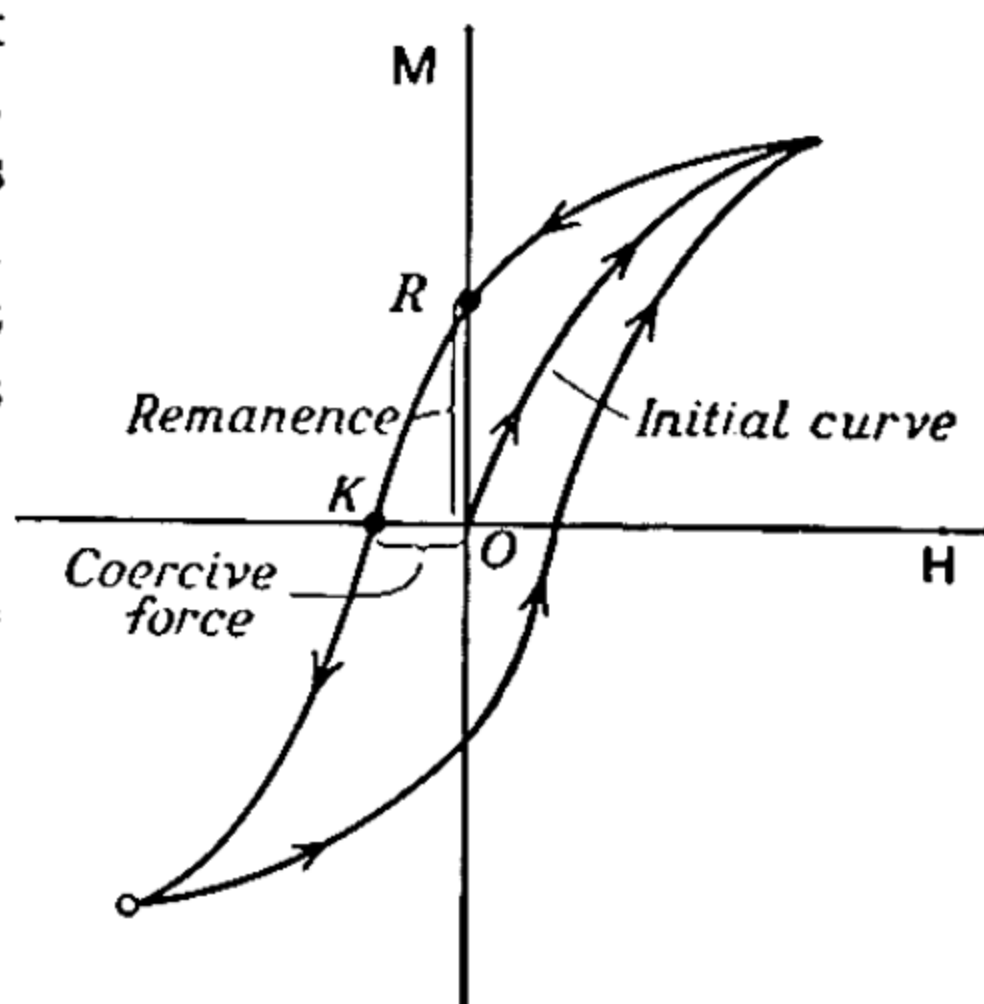


Fig. 2

Even with the origin of the constant  $\nu$  explained, there are still many difficulties connected with an understanding of the actual magnetization curve. It is well known that this curve has the form of a loop called the *hysteresis loop* (fig. 2). It follows that very different values of  $M$  correspond to a given value of  $H$ , depending on the previous treatment of the sample. One of the first questions arising is: At very small field values, why does not the magnetization jump to one of the two points of intersection of the curve and straight line in fig. 2? Considerations of stability show that these points represent stable states, while the condition of zero magnetization is an unstable one. In reality, the spontaneous magnetization deduced from fig. 1 corresponds to the condition of *saturation* attained with the highest fields available. The external fields are so small in comparison with the internal one that they would produce only tiny displacements of the straight lines in fig. 1 when represented on the correct scale. These

external fields obviously serve only to overcome certain constraints that limit the value of the magnetization, which is large even for zero field. Inasmuch as these constraints depend strongly on the material and on its previous treatment, a great variety of magnetization curves is obtained.

The saturation condition may be changed by lowering the temperature, which corresponds to rotation of the straight lines. Measurements of the temperature-dependence of the spontaneous magnetization of iron and nickel are in accord with theory if the  $L$  curve in fig. 1 is represented by  $\tanh x$ . As will be seen later (p. 667), together with the gyromagnetic anomaly factor 2 discussed in the next section, this means that the electrons responsible for ferromagnetism have no orbital moment. However, it is not correct to conclude that these are conduction electrons, for in the salts of the entire iron group the mutual disturbance of the electron orbits makes these substances behave magnetically as though the electrons were free (see p. 669 *et seq.*).

The question raised above, as to why a considerable field is needed to produce spontaneous magnetization, is to a great extent answered by investigations on single crystals. The curves for various directions of magnetization in iron are given in fig. 3. These curves may be understood on the basis of the following considerations: There are apparently preferred directions of magnetization in a crystal. For the cubic lattice of iron, for example, this is the  $[100]$  direction—that of the edges of the cube. No distinction exists between a given direction and its opposite. Apparently at zero field there are sizable regions or *domains* oriented in one of the six preferred directions, but because of the compensating effects of opposite senses no external magnetic effects are noticed. When the field is applied, the domains snap into the field direction to the extent that it is along an edge. Since the directions are energetically equal, no expenditure of energy is required and the lining-up proceeds very readily as shown by the steepness of the curve for this direction. Of course, there may be a potential barrier between the two opposite directions, i.e. the lining-up process proceeds along directions corresponding to greater energy. As a result, the magnetization curves are not strictly vertical at the start; also, many domains remain in their former orientation after removal of the field because the thermal energy is not large enough to surmount the potential barrier. This is the way remanence is explained. If the field is in a different direction—for instance, that of the cube diagonals  $[111]$ —all domains line up in the directions of the edges lying nearest the field direction (fig. 4). Further increase of the resultant magnetization can occur only by the turning of domains away from the directions of the edges, and this requires energy. As a result, the process of magnetization goes easily up to  $M_s/\sqrt{3}$ , but beyond this point the curve must rise much less steeply. That this really is the case is shown by



fig. 3. Upon removal of the field the rotated domains will promptly go back to the edge directions, which correspond to minimum potential energy. This means that the turning process, unlike the reversal process, is free of hysteresis. For fields in the direction of the face diagonal the curve begins to bend over at  $M_s/\sqrt{2}$ .

For technical polycrystalline materials the actual curve of magnetization is not given simply by superposition of the various crystal effects. Rather, an additional circumstance—internal stress—exists, and this also determines a preferred direction. The preferential direction in a stretched iron wire is that of the tension. In a region of

uniform internal stress the actual magnetization vector orients itself in such way that the sum of the field energy, crystal energy and tensile energy is a minimum. Since the stresses vary from place to place, neighbouring domains differ in

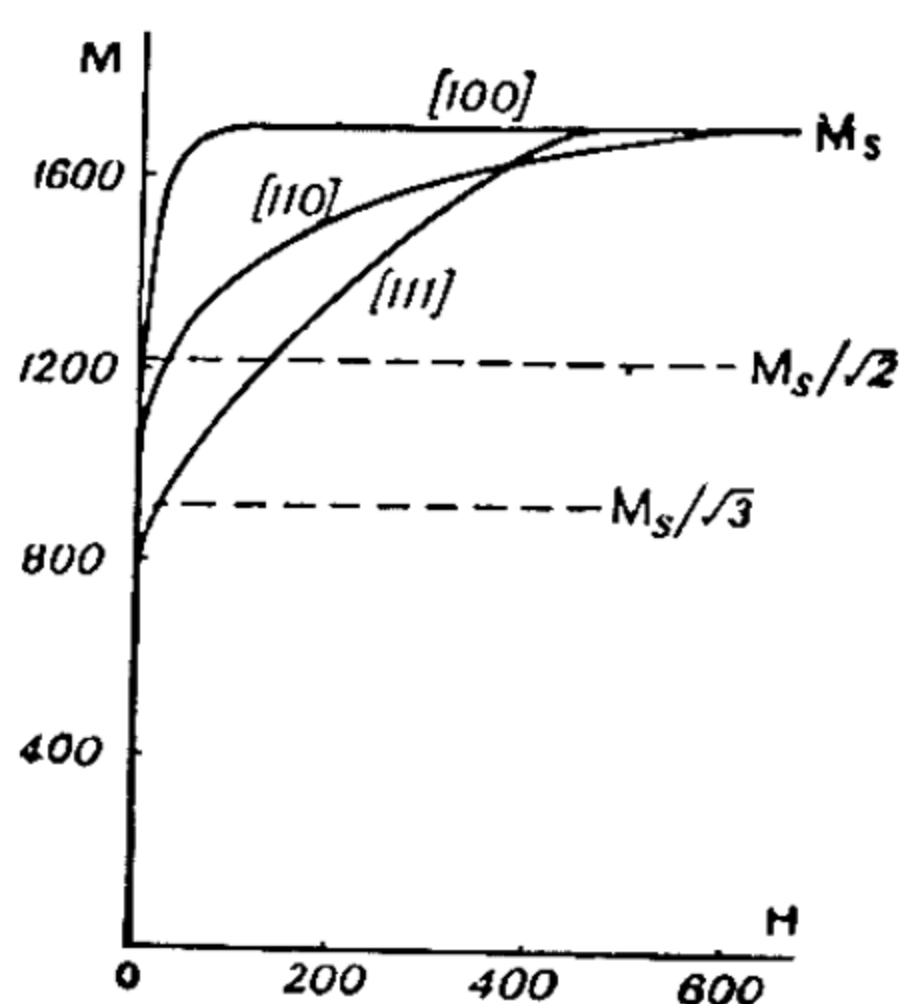


Fig. 3

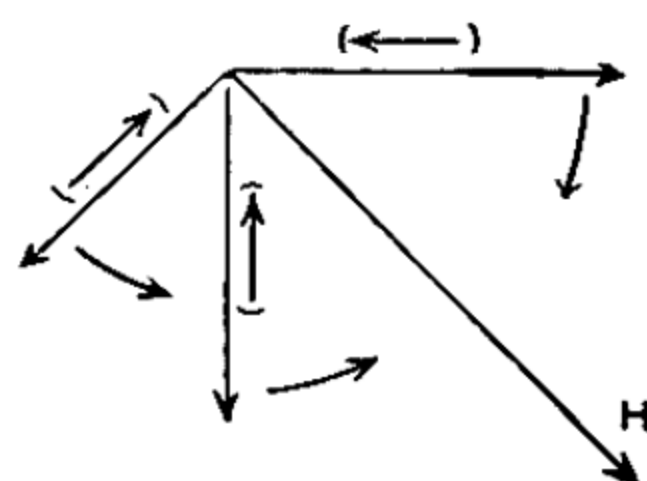


Fig. 4

energy and the external field has the additional effect of making individual domains grow at the expense of others. Quantitative calculation of these ideas is quite troublesome and the details will not be given here. It is sufficient to note that such considerations have in recent years led to an almost complete understanding of what seemed at one time to be a hopelessly confused subject.

## 5. Magnetism Induced by Revolving Electrons. The Magnetomechanical Parallelism. Theory of Diamagnetic Susceptibility.

If we inquire into the cause of the magnetic moment of an atom, basing our considerations on our present concepts of atomic structure, we arrive at a vivid interpretation of the "elementary currents" which Ampère assumed were the origin of atomic magnetism. According to our modern point of view, these currents are to be identified with the orbital motion of electrons about the nucleus. If the period of revolution of an electron is  $\tau$  sec., its orbit is equivalent to a circular



current of strength  $-e/\tau$ . Taking the orbit to be plane and calling the enclosed area  $\mathbf{S}$ , the magnetic moment becomes

$$\boldsymbol{\phi} = -\frac{e}{c} \cdot \frac{\mathbf{S}}{\tau} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (37)$$

Since the motion of the electron takes place under a central force, the Law of Areas (p. 89) is valid, and we may replace  $\mathbf{S}/\tau$  by  $d\mathbf{S}/dt$ . Now the areal velocity is given in direction and magnitude by the mechanical moment of momentum  $\boldsymbol{d}$  divided by  $2m$ ; we then have for the magnetic moment

$$\boldsymbol{\phi} = -\frac{e}{2mc} \boldsymbol{d} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (38)$$

This important relationship is called the *magnetomechanical parallelism*. To every magnetic moment there corresponds a mechanical moment of momentum whose existence is due to the inertial mass of the electrons. This can be experimentally verified. Since the conservation of angular momentum requires this quantity to remain constant when there are no external forces, a change in the magnetic state of a body must impart to it a rotational motion whose angular momentum is equal and opposite to that of the electrons. This phenomenon is known as the Richardson-Einstein-de Haas Effect, and is actually observed. The magnetic state of the body may be altered by magnetizing it or by destroying its magnetism by heating it above the Curie point. Formula (38) may be checked by observing the values of magnetic moment and moment of momentum; surprisingly enough, the factor on the right comes out double the value  $e/2mc$  for all ferromagnetic substances; for salts of the rare earths, for which the effect is so small that it is just measurable, a value between one and two times this factor results. We shall take up this significant anomaly in detail later (p. 667).

The inverse effect—magnetization by rotation—was first observed by S. J. Barnett in 1914, shortly before the discovery of the direct effect.

If an atom has several electrons, it is possible for the mechanical angular momenta—and hence the magnetic moments—to mutually compensate each other. In such an atom, the field gives rise to a moment opposite in direction to the field; its magnitude will be calculated immediately. Naturally, this induced moment appears also in paramagnetic atoms, but is small compared with the moment already existing, and so may be neglected in general. According to the theorem of Larmor mentioned above and derived on p. 664, the magnetic field brings about an added rotation of the entire electron system around

an axis through the nucleus parallel to  $\mathbf{H}$ . The angular velocity of this motion is given in direction as well as in magnitude by

$$\boldsymbol{\omega} = \frac{e}{2mc} \mathbf{H}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (39)$$

This rotation gives rise to an additional angular momentum

$$\mathbf{d} = m \sum [\mathbf{r}_i \mathbf{v}_i] = m \sum [\mathbf{r}_i [\boldsymbol{\omega} \mathbf{r}_i]] = m \sum (\boldsymbol{\omega} \cdot \mathbf{r}_i^2 - \mathbf{r}_i \cdot \mathbf{r}_i \boldsymbol{\omega}). \quad (40)$$

The summation is to be extended over all electrons. The radius vector of the  $i$ th electron, referred to the nucleus, is  $\mathbf{r}_i$ . Take the  $z$ -axis of a rectangular system of co-ordinates in the direction of the field. Then the  $z$ -component of this angular momentum, which is the only component having a non-vanishing resultant, becomes

$$d_z = m \omega \sum (r_i^2 - r_i^2 \cos^2 \theta_i) = m \omega \sum (x_i^2 + y_i^2). \quad . \quad (41)$$

Since the field has no aligning effect, all orientations of the atom are, on the average, equally probable, and since  $x^2 + y^2 + z^2 = r^2$ ,

$$\overline{x_i^2} = \overline{y_i^2} = \overline{z_i^2} = \frac{\overline{r_i^2}}{3}. \quad . \quad . \quad . \quad . \quad . \quad (42)$$

Hence by (38) the average induced magnetic moment is

$$\bar{p} = \bar{p}_z = -\frac{e\omega}{3c} \sum r_i^2 = -\frac{e^2 \sum \overline{r_i^2}}{6mc^2} H, \quad . \quad . \quad . \quad (43)$$

and the molar susceptibility is then given by

$$\chi_m = -\frac{e^2 L}{6mc^2} \sum \overline{r_i^2}. \quad . \quad . \quad . \quad . \quad . \quad (44)$$

As one sees, the negative sign of the diamagnetic susceptibility results correctly; in addition, the mean orbital radii computed from (44) agree well with the values obtained from other data.

## CHAPTER XXVI

### PHENOMENOLOGICAL THEORY OF SUPERCONDUCTIVITY

#### 1. The Fundamental Equations.

The superconducting state is characterized chiefly by the sudden drop in the electrical resistance of a conductor when it is cooled to temperatures near the absolute zero. This effect was discovered by H. Kammerlingh Onnes in 1911. The decline from the finite resistance corresponding to the temperature of the conductor to the immeasurably small value characteristic of this phenomenon takes place in an interval of only a few hundredths of a degree. The midpoint of this interval is called the critical transition temperature. For pure metals this lies between 0.1 and 10° K., while for metallic conducting compounds such as hydrides and nitrides of metals it falls somewhat higher. Since new superconductors are continually being discovered, it is feasible to state only the groups of elements among which no indication of superconductivity has yet been found. These are the elements of the first column of the periodic table (the alkali metals and Cu, Ag and Au) and the ferromagnetic metals.

The vanishing of the electrical resistance is not the only characteristic of superconductivity. A second circumstance, discovered by Meissner and Ochsenfeld, may be roughly described as the vanishing of the permeability. If a sample which is already in the superconducting state is brought into a magnetic field, current will be induced in the surface layers in such manner as to cancel the field within the sample. The external magnetic field behaves as if the material had zero permeability. However, if the sample is cooled below the transition temperature while in the magnetic field, nothing should occur according to Maxwell's theory. Meissner found, nevertheless, that exactly the same field distribution set in as in the first instance. One might expect to deduce the behaviour of superconductors by setting  $\sigma = \infty$  and  $\mu = 0$  in Maxwell's equations. This does not work.

Although no satisfactory model has yet been suggested, London and later von Laue found a way to supplement the Maxwell equations so as to represent correctly the phenomena and to describe them in terms of a single material constant which depends on the temperature. For the present, the Maxwell equations will be written as they stand, with  $\epsilon$  set



equal to unity for metals. We retain  $\mu$  for the permeability at the transition temperature, having in mind the fact that certain phenomena have not been completely explained, although the most recent results indicate that  $\mu$  may always be taken equal to unity.

The Maxwell equations may now be written

$$\text{curl } \mathbf{H} = \frac{1}{c} \frac{\partial \mathbf{E}}{\partial t} + \frac{4\pi \mathbf{i}}{c}. \quad . \quad . \quad . \quad . \quad . \quad (1)$$

$$\text{curl } \mathbf{E} = -\frac{1}{c} \frac{\partial \mathbf{B}}{\partial t}. \quad . \quad . \quad . \quad . \quad . \quad (2)$$

$$\text{div } \mathbf{E} = 4\pi \rho. \quad . \quad . \quad . \quad . \quad . \quad (3)$$

$$\text{div } \mathbf{B} = 0. \quad . \quad . \quad . \quad . \quad . \quad (4)$$

$$\mathbf{B} = \mu \mathbf{H}. \quad . \quad . \quad . \quad . \quad . \quad (5)$$

The new feature is that  $\mathbf{i}$  and  $\rho$  are each to be broken down into a normal and a "super" part. The behaviour of superconductors in high-frequency fields shows that there is an ohmic contribution to the current at temperatures below as well as above the transition temperature. It appears that only a portion of the electron ensemble participates in superconduction, while the remainder behaves in normal fashion. Accordingly we take

$$\mathbf{i} = \mathbf{i}_n + \mathbf{i}_s; \quad \rho = \rho_n + \rho_s.$$

The following equation of continuity holds for each part independently:

$$\text{div } \mathbf{i}_n + \frac{\partial \rho_n}{\partial t} = 0; \quad \text{div } \mathbf{i}_s + \frac{\partial \rho_s}{\partial t} = 0. \quad . \quad . \quad . \quad (6)$$

As before, we put for the normal current

$$\mathbf{i}_n = \sigma \mathbf{E}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

From the fact that for an electron subject to no "frictional" resistance the acceleration will be proportional to the field intensity, London assumed

$$\frac{\partial}{\partial t} (\lambda \mathbf{i}_s) = \mathbf{E}, \quad . \quad . \quad . \quad . \quad . \quad (8)$$

where  $\lambda$  represents a new constant of the material which has a value dependent upon the temperature.

For the relation of the super-current and the magnetic induction (field intensity), a relation is set up which, in principle, amounts to



Taking the curl of (12) and using (13') and remembering that  $\text{div } \mu \mathbf{H} = 0$  and  $\text{div } \mathbf{i} = 0$ , we get

$$\text{curl curl } \mathbf{H} = \frac{4\pi}{c} \text{curl } \mathbf{i} = -\frac{4\pi\mu}{c^2\lambda} \mathbf{H}. \quad (14)$$

Similarly,

$$\text{curl curl } \mathbf{i} = -\frac{\mu}{c\lambda} \text{curl } \mathbf{H} = -\frac{4\pi\mu}{c^2\lambda} \mathbf{i}. \quad (15)$$

Imagine a superconductor of infinite extent bounded by the plane  $z = 0$  and filling all space on the positive side of the  $z$ -axis. Let the space where  $z < 0$  be a vacuum where there exists a constant magnetic field  $\mathbf{H}_0$ . By virtue of (10) and because

$$\text{div } \mathbf{H} = \frac{\partial H_x}{\partial x} + \frac{\partial H_y}{\partial y} + \frac{\partial H_z}{\partial z} = 0$$

$H_z$  must be constant in both regions. But  $H_z = \text{const.}$  is not a solution of (14) except for  $H_z = 0$ . This means that  $\mathbf{H}$  can have only a tangential component, which we take in the  $y$ -direction. Thus

$$\text{For } z < 0, \quad H_z = H_x = 0; \quad H_y = H_0.$$

$$\text{For } z > 0, \quad H_z = H_x = 0; \quad H_y = f(z).$$

Because of the continuity of  $\mathbf{H}$ , and by virtue of (14), we shall have, in the superconductor

$$H_y = H_0 e^{-\beta z}, \quad \text{where } \beta = \frac{1}{c} \sqrt{\frac{4\pi\mu}{\lambda}}. \quad (16)$$

From (12), the current density is given by

$$i_x = -\frac{c}{4\pi} \frac{\partial H_y}{\partial z} = -\frac{\beta c}{4\pi} H_0 e^{-\beta z}; \quad i_y = 0; \quad i_z = 0. \quad (17)$$

In a superconductor, then, the current as well as the magnetic field fall off with the depth. The penetration depth  $1/\beta$  is proportional to  $\sqrt{\lambda}$ . Independent measurements always lead to an order of magnitude of about  $10^{-5}$  cm. for  $1/\beta$  for a temperature of  $0.5^\circ$  below the transition point. The Meissner effect is thus accounted for.

### 3. Optical Behaviour of Superconductors.

In high-frequency (optical) fields, the ratio of super- to normal current is as follows: Because of the connexion of the two through the electrical field intensity  $\mathbf{E}$  (equations (7) and (8)) we always have

$$\mathbf{i}_n = \sigma \frac{\partial}{\partial t} (\lambda \mathbf{i}_s). \quad (18)$$



If  $i_s$  is a periodic function of the time,

$$i_s = i_{s,0} e^{i\omega t}, \quad . . . . . (19)$$

then

$$i_n = \sigma\omega\lambda i_s, e^{i\omega(t+\pi/2)}. \quad . . . . . (20)$$

The normal current leads the super-current by a phase angle  $\pi/2$ . The amplitude ratio is given by the product  $\sigma\omega\lambda$ . Taking for the order of magnitude of  $\lambda$  the depth of penetration mentioned above, the amplitude of the normal current turns out to be a hundred times that of the super-current even for the infra-red region. This is in harmony with the negative result of all experimental attempts to detect optical changes when superconduction sets in. This circumstance compels us to assume that a normal current exists in addition to the super-current.

Although the Laue-London theory embraces a complete picture of the phenomena, elaborate computations have not yet succeeded in supplying an atomistic theory of the superconductivity constant  $\lambda$  along lines similar to our interpretation of the material constants  $\sigma$ ,  $\epsilon$ , and  $\mu$  as presented in the previous sections. To this extent, the present topic still belongs to the phenomenological part of electricity. However, it was presented here for two reasons: Firstly, certain atomistic notions, such as the acceleration of electrons, were used in setting up the field equations; secondly, there is the hope that an addendum concerned with an atomistic verification may be forthcoming as the subject develops.

## CHAPTER XXVII

### THE ELECTRODYNAMICS OF MOVING BODIES

#### 1. Electromagnetic Induction in moving Bodies from the Standpoint of the Electron Theory.

An exact formulation of *all* electromagnetic phenomena in moving bodies is possible only with the aid of the Theory of Relativity; in fact, consideration of these phenomena formed the starting-point of this theory. Nevertheless, the Electron Theory is of such great service in representing most of the phenomena that, for the sake of simplicity of deduction, it seems advisable to consider matters from this standpoint first, rather than to begin at once to attack these problems with the heavy ordnance of the mathematics of the relativity theory. This is particularly so in view of the fact that up to the present time only a few experiments are known which are sufficiently accurate to detect the small differences in the results of the two theories (p. 476). The Electron Theory assumes that the electromagnetic field is located in a stationary "ether", i.e. in an absolute space pervading all matter, while charge, polarization and magnetization are bound to the material bodies and share in their motions. In our consideration of relativistic mechanics we pointed out that the existence of an all-pervading ether is disproved by the results of three experiments whose accuracy is sufficient to settle the matter. As a consequence, all results of the following calculations are merely approximations, but very useful ones nevertheless. If we write the field equations in the form \*

$$\text{curl } \mathbf{H} = \frac{1}{c} \frac{\partial \mathbf{E}}{\partial t} + \frac{4\pi}{c} \frac{\partial \mathbf{P}}{\partial t} + \frac{4\pi}{c} (\rho_+ \mathbf{v}_+ - \rho_- \mathbf{v}_-), \quad . \quad (1)$$

$$\text{curl } \mathbf{E} = -\frac{1}{c} \frac{\partial \mathbf{H}}{\partial t} - \frac{4\pi}{c} \frac{\partial \mathbf{M}}{\partial t}, \quad . \quad . \quad . \quad (2)$$

$$\text{div } \mathbf{E} = 4\pi(\rho_+ - \rho_-), \quad . \quad . \quad . \quad (3)$$

$$\text{div } \mathbf{H} = 0, \quad . \quad . \quad . \quad (4)$$

\* The formulation used here differs from the usual one in which the conduction term is simply  $(4\pi/c)\rho\mathbf{v}$ . The present form is chosen for the reason that, for example, in electrolytes there is a conduction current even when the space charge in (3) is not different from zero. In that case, however, equations (1) and (3) written in the usual manner would contradict each other.

the terms  $\partial \mathbf{P}/\partial t$ ,  $\partial \mathbf{M}/\partial t$ ,  $\rho$  and  $\rho \mathbf{v}$  are affected by the motions of material bodies. In addition to these equations we have the formula for the force exerted by an electromagnetic field on a moving charge  $e$ :

$$\mathbf{F} = e\mathbf{E} + \frac{e}{c} [\mathbf{v}\mathbf{H}]. \quad . . . . . (5)$$

Thus, while a magnetic field exerts no force on a charge at rest, a force comes into being as soon as the charge is set in motion. If we look upon every force acting on an electrical charge as an electrical force, we have an "induced electrical field strength"

$$\mathbf{E}' = \frac{1}{c} [\mathbf{v}\mathbf{H}]. \quad . . . . . (5')$$

We now consider a variety of experiments from this point of view.

(a) *Ordinary induction in a moving coil*

The Law of Induction embodied in the second Maxwell equation is based upon the assumption that the change in flux consists of a time variation of the magnetic field, the path of integration being held fixed. Nevertheless, the law of induction is also applied without further thought to the case in which the flux is altered by moving the conductor while the field remains constant. With the help of equation (4) we shall show that this is really justified. If the induced electric field in (5') is integrated along a conducting loop there results

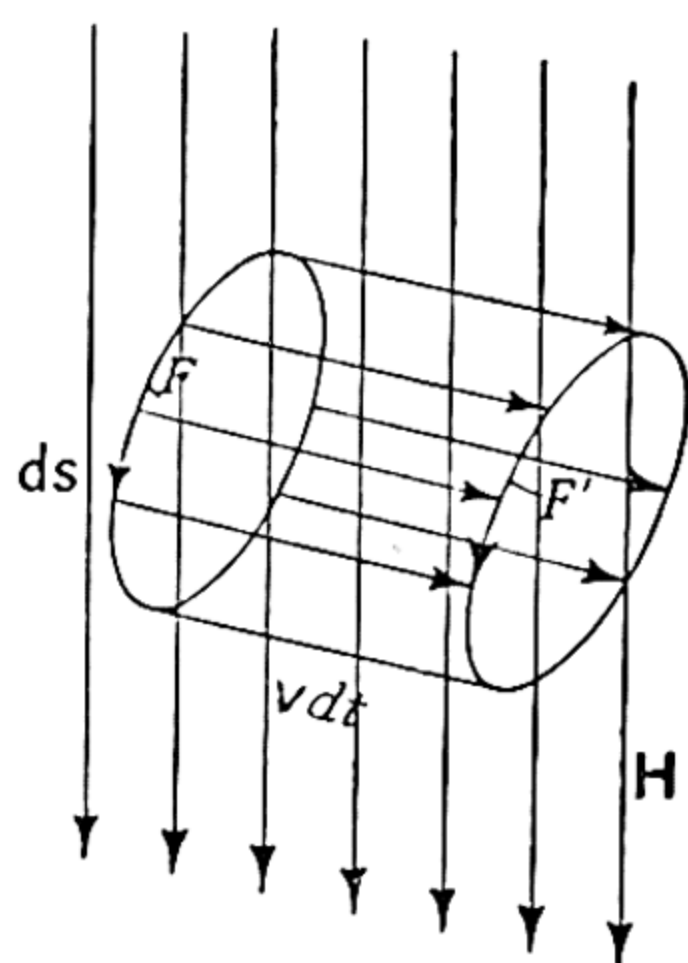


Fig. 1

$$\oint \mathbf{E}' d\mathbf{s} = \frac{1}{c} \oint [\mathbf{v}\mathbf{H}] d\mathbf{s}. \quad . (6)$$

Now it may be shown readily that the right member is equal to the rate of change of the magnetic flux accompanying the motion of the conductor. Let us consider two neighbouring positions of the loop at times  $t$  and  $t + dt$ . The change in position may also involve a deformation. Connect corresponding points of the two contours by displacement vectors  $\mathbf{v} dt$ . We thus complete a closed surface over which we take the integral of the magnetic field strength. Since the divergence of  $\mathbf{H}$  vanishes in a uniform medium of permeability  $\mu$ , this surface integral is zero, by Gauss's theorem. Now the element of the lateral surface (see fig. 1) may be expressed as

$$d\mathbf{S}_1 = - [\mathbf{v} d\mathbf{s}] dt,$$



$$\text{so that} \quad 0 = - \int_r \mathbf{H} d\mathbf{S} + \int_r \mathbf{H} d\mathbf{S} - dt \int_{\text{side}} \mathbf{H} [v d\mathbf{S}].$$

$$\text{Hence} \quad d\Phi = \int_r \mathbf{H} d\mathbf{S} - \int_r \mathbf{H} d\mathbf{S} = - dt \int_{\text{side}} [v \mathbf{H}] d\mathbf{S}, \quad . \quad . \quad (7)$$

$$\text{i.e.} \quad \int [v \mathbf{H}] d\mathbf{S} = - \frac{d\Phi}{dt}. \quad . \quad . \quad . \quad . \quad . \quad (8)$$

In a medium of permeability  $\mu$  the force on the moving charge is  $\mu$  times as large, so that we obtain  $\mu$  times the induced potential and the integral in (8) represents the flux of *induction*. Thus it amounts to the same thing if the change in time of the induction flux takes place through a change of field for a fixed loop, or by the motion of the loop in a constant field—a fact tacitly accepted in electrical engineering.

(b) *Dielectric polarization induced in insulators*

H. A. Wilson performed the experiment of rotating a hollow cylinder of a dielectric substance in a magnetic field parallel to the axis. The induced electric field acting on the quasi-elastically bound electrons of the substance displaces them relative to the nuclei and causes a resultant polarization. Neglecting the internal field, this quantity is given by

$$\mathbf{P} = \frac{N\alpha[v\mathbf{H}]}{c}, \quad . \quad . \quad . \quad . \quad . \quad (9)$$

where  $N$  is the number of atoms per unit volume and  $\alpha$  their polarizability. Using the relationship between  $N\alpha$  and  $K$  (cf. p. 452), we obtain

$$\mathbf{P} = \frac{K-1}{4\pi} \frac{[v\mathbf{H}]}{c}. \quad . \quad . \quad . \quad . \quad . \quad (10)$$

In our case  $\mathbf{P}$  is radial and has the magnitude

$$P = \frac{(K-1)vH}{4\pi c}. \quad . \quad . \quad . \quad . \quad . \quad (10')$$

The rotating dielectric has thus assumed a volume polarization and is to be compared with a permanently polarized body or "electret". There are free charges on the surface, their surface density  $\sigma$  being given by the magnitude of  $\mathbf{P}$ , according to p. 281. The detection of these charges is accomplished as follows: the inner and outer surfaces of the cylinder are covered with metal. For the present case it is immaterial whether the metal coatings rotate along with the cylinder or not, for the magnetic effects due to the moving charges are not measured here. If, during the rotation, the two metal coverings are connected and earthed, the total flux emanating from the surface of the dielectric ends on the coatings, i.e. there is induced upon them a charge equal and opposite to the free charge on the surface of the dielectric. If we now break the connexion between the coatings and

annul the volume polarization of the dielectric by stopping the rotation or cutting off the magnetic field, a surface charge of density  $\sigma = P$  remains on the coatings. This charge gives rise to a potential difference between the coatings which can be measured by an electrometer. This difference of potential depends on the capacity of the cylindrical condenser and of the electrometer. The measurements confirm the predictions of the theory.

(c) *W. Wien's experiment*

The displacement of charge (i.e. polarization) generated in atoms by motion in a magnetic field may be detected in a particularly favourable optical case. As discovered by J. Stark, the spectral lines of hydrogen are split up into numerous components in an electric field. This is the so-called Stark effect. Wien caused hydrogen canal rays to pass through a magnetic field which was normal to their direction of motion. This is equivalent to an electrical field normal to the plane of  $\mathbf{v}$  and  $\mathbf{H}$  and of magnitude  $vH/c$ . By observing in the direction of the magnetic field one actually sees a pattern identical with the Stark effect pattern obtained in transverse observation in an electric field.

## 2. Magnetic Effects of Moving Charges.

(a) *Rowland's experiment*

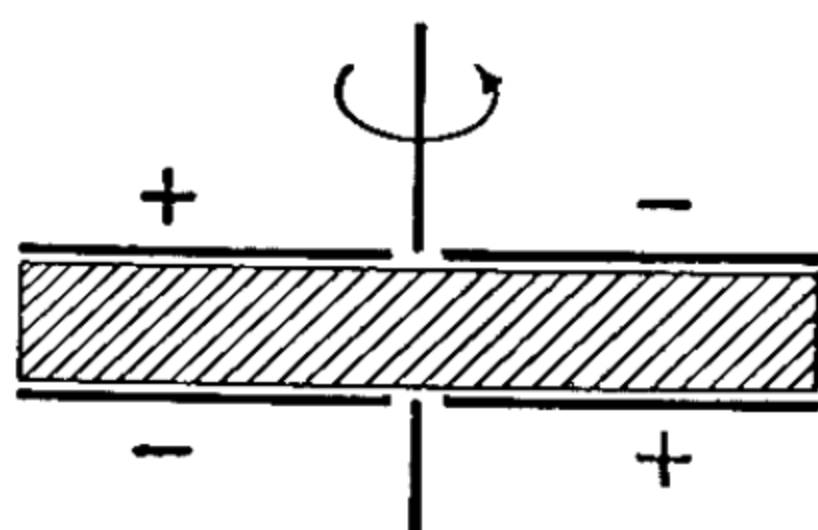
The next experiment which suggests itself is the testing of the equivalence of a mechanically moved charge with an electric current. This is accomplished by imparting a large velocity to a charged body and detecting the magnetic effects of this artificial rotational current. H. A. Rowland set a charged, insulated metal plate into rapid rotation and obtained the magnetic effect given by the third term of the right member of equation (1) (p. 471).

(b) *The experiments of Röntgen and Eichenwald*

Röntgen and Eichenwald were able to show that the charges residing on the surfaces of a polarized dielectric ( $\sigma = |\mathbf{P}|$ ) correspond to a current when the body is set into motion. To demonstrate this, Röntgen caused the dielectric between the plates of a charged condenser to rotate; the magnetic effect calculated from the surface density of charge and the rotational speed was found to agree with the observed value.

Using another arrangement, the same investigators were able to show that the polarization could be altered by mechanical motion; this change causes a magnetic field corresponding to the second term of equation (1) (p. 471). In this experiment we again make use of a rotating dielectric disc, but the condenser plates are divided into two

halves along a diameter, and adjacent parts are held at opposite potentials (fig. 2). Thus the sign of the polarization at any point reverses when this portion crosses the line of separation, so that a vertical "displacement current" flows in the separating layer. The existence of this current was detected by its magnetic effects.

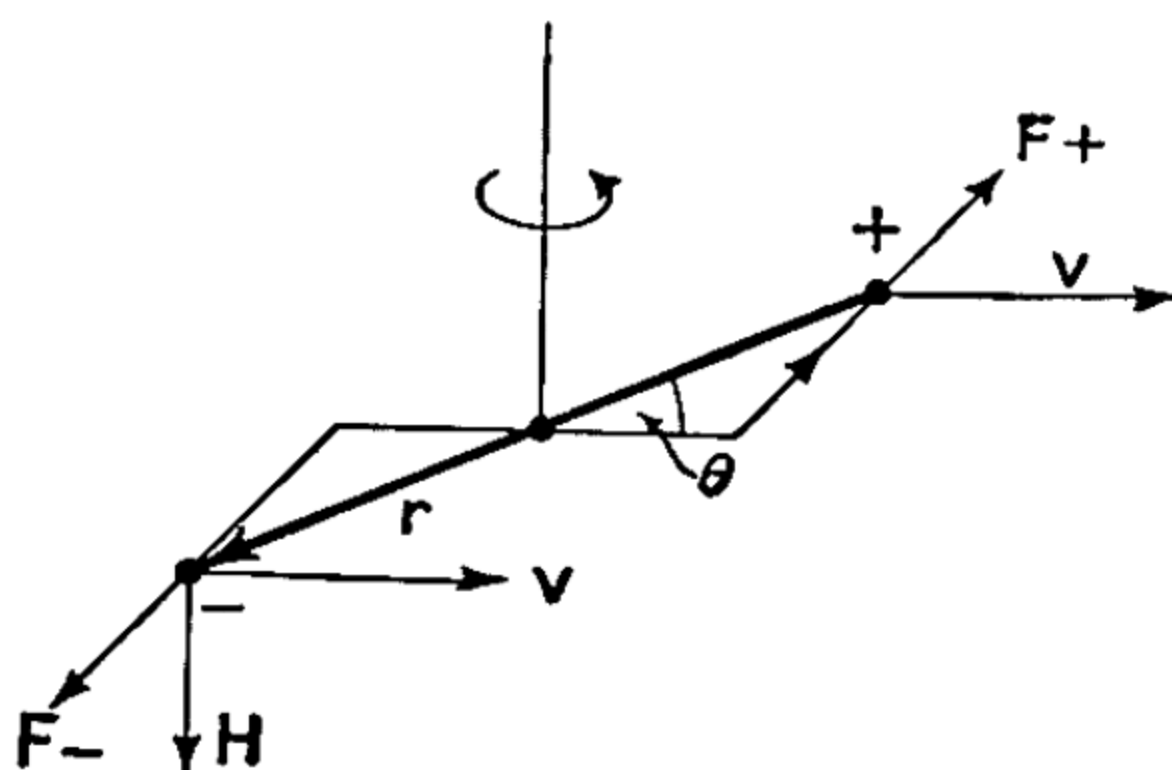


**Fig. 2**

(c) *The experiment of Trouton and Noble*

Trouton and Noble suspended a parallel plate condenser in such manner that it could turn freely about an axis parallel to the plates.\*

The effect to be expected is that, as a result of the translatory motion of the earth, the charged condenser will turn into a position in which the plates are parallel to the direction of this motion. In order to see that, if we use the ether concept, the motion of the earth would cause a torque, it is convenient to imagine the plate charges  $e$  concentrated in two small



**Fig. 3**

spheres whose separation is equal to that of the condenser plates (fig. 3). Let the density of charge in these spheres be  $\rho$ . The translatory motion of the charge  $+e$  due to the earth's motion corresponds to a current element  $ds$  for which

$$IdS = ev.\dagger \quad . \quad . \quad . \quad . \quad . \quad . \quad (11)$$

According to the Biot-Savart Law, this current element gives rise to a magnetic field at the position of the negative charge. This field is given by

$$\mathbf{H} = \frac{I[ds\mathbf{r}]}{cr^3} = e \frac{[\mathbf{v}\mathbf{r}]}{cr^3}, \quad \cdot \cdot \cdot \cdot \cdot \quad (12)$$

where  $\mathbf{r}$  is the vector drawn from the positive to the negative charge. Thus the field is normal to the plane of  $\mathbf{v}$  and  $\mathbf{r}$ . The force with which

\* F. Trouton and T. Noble, *Phil. Trans.*, A, 202, p. 165 (1903); *Proc. Roy. Soc.*, 72, p. 132 (1903).

† If we imagine the sphere divided into slices normal to  $v$ , each of thickness  $ds$ , then each slice corresponds to a current element  $\rho A ds \cdot v$ . Integration over all these slices yields the volume of the sphere multiplied by  $\rho v$ , whence (11) follows.



this field acts on the negative charge, which is also moving with velocity  $\mathbf{v}$ , is given by

$$\mathbf{F} = -e \frac{[\mathbf{v}\mathbf{H}]}{c} = -\frac{e^2}{c^2} \frac{[\mathbf{v}[\mathbf{v}\mathbf{r}]]}{r^3}. \quad \dots \quad (13)$$

This force, which is normal to the plane of  $\mathbf{H}$  and  $\mathbf{v}$ , is of magnitude

$$F = \frac{e^2 v^2 \sin \theta}{c^2 r^2}. \quad \dots \quad (14)$$

Similarly, an equal and opposite force acts upon the positive charge as a result of the field caused by the motion of the negative charge. The two forces together furnish a torque parallel to the axis of rotation and of magnitude

$$\begin{aligned} M &= Fr \cos \theta = \frac{e^2 v^2}{r c^2} \sin \theta \cos \theta \\ &= \frac{1}{2} \frac{e^2 v^2}{r c^2} \sin 2\theta. \quad \dots \quad (15) \end{aligned}$$

This turning moment tends to turn the plates parallel to the "ether wind". Now  $e^2/r$  is the electrostatic energy  $U_{el}$  of the system. Inserting this value in equation (15) we obtain the correct formula for the parallel plate condenser, except for a factor  $1/2$ , which is due to the simplifying assumptions made regarding the distribution of charge. The correct formula is

$$M = U_{el} \left( \frac{v}{c} \right)^2 \sin 2\theta. \quad \dots \quad (16)$$

Thus a condenser oriented at an angle  $\pi/4$  with the direction of the earth's orbit should be subject to a torque  $U_{el}(v/c)^2$  when charged. A moment of this magnitude should be easily observable, yet even the most delicate experiments, especially those conducted by Tomaschek, failed to give any indication of such a torque. *The result of the Trouton-Noble experiment, which is one of the very few experiments permitting of observation of quantities of the second order in  $v/c$ , contradicts the theory of a stationary ether.*

#### (d) Variability of the mass of the electron

The magnetic field of a moving point charge represents a certain energy distributed in space, the generation of which requires a certain amount of work; i.e. the electron possesses inertia which, as calculation shows, increases with increasing velocity. But since the mass is manifested only by its inertia, this indicates that the inertial mass of an electron increases with velocity. The amount of this increase is different for acceleration in the direction of motion and normal to this direction. We therefore speak of *longitudinal mass* and *transverse mass*.

According to a calculation made by Abraham, these quantities for a rigid spherical electron are:

$$\text{longitudinal mass } m_l = m_0 \left( 1 + \frac{6}{5} \frac{v^2}{c^2} \right), \quad . \quad . \quad (17)$$

$$\text{transverse mass } m_t = m_0 \left( 1 + \frac{6}{15} \frac{v^2}{c^2} \right). \quad . \quad . \quad (18)$$

Again, the lowest power of  $v/c$  which appears is the *second*. In addition to our previous experiments this phenomenon may also be called upon to decide for or against the concept of a stationary ether. As mentioned on p. 430, this variability of electron mass has been observed for some time. The theory of relativity leads to a simpler formula which also contains correction terms in  $(v/c)^2$  and which deviates from the above by about twenty per cent. *The most exact measurements have decided unanimously against the formulæ (17) and (18) derived from the ether concept, and in favour of the relativistic formula.*

### 3. Propagation of Electromagnetic Waves in Moving Media.

#### (a) Fizeau's experiment

The classical experiment on the propagation of light in moving media is that of Fizeau \* in which light was made to pass through flowing water. The small difference between the index of refraction of moving and stationary water was measured by an interference method. Approaching the question quite naïvely, one might believe that the velocity of the flowing water simply would be added to the phase velocity as given by the index of refraction for water at rest. It is apparent, however, that this view leads one into difficulties, for evidently this convection effect must exist to the full extent even for very tenuous moving matter (e.g. in a rapidly flowing gas), while in a perfect vacuum there can be no effect. This would mean a discontinuity contrary to all experience. The actual result of the experiment showed that the wave velocity  $u$  is increased by the velocity of flow  $v$  multiplied by the factor  $(1 - 1/n^2)$ . The factor  $(1 - 1/n^2)$  which approaches zero when  $n$  approaches unity is called Fresnel's convection coefficient. This result is obtainable also from the theory of a stationary ether, but follows much more simply from the relativity theory as a consequence of the addition theorem for velocities (cf. p. 244).

#### (b) The Michelson-Morley experiment

A large number of experiments intended to detect the influence of the motion of the earth on the propagation of light have been proposed. For all of these, with one exception—the Michelson-Morley experiment—the accuracy of observation is not great enough to detect

\* H. Fizeau, *Comptes rendus*, 33, p. 349 (1851); *Ann. d. Phys. u. Chem.*, Erg., 8, p. 457 (1853).



deviations of higher order than the first power of  $v/c$ . The stationary ether theory yields no first-order effect for a common uniform translation of source, observer and intervening parts of the apparatus. To that extent all these experiments also confirm the theory of a stationary ether. This theory, however, does demand a second-order effect, and the Michelson-Morley experiment is the only optical experiment sufficiently sensitive to detect this effect. This experiment has already been discussed in detail on p. 238. *The absence of a second-order effect here, the negative result of the Trouton-Noble experiment and the departures of the mass variability of the electron from the Abraham formula (p. 477) are evidence for the fact that the concept of a stationary ether cannot correspond with reality.*

(c) *Doppler effect and aberration of light*

These phenomena have been discussed in detail on p. 245.

(d) *The experiment of Sagnac*

In this experiment the source of light, the observer, and the intermediate apparatus are given a common *rotational* motion.\* Thus, in spite of the fact that all parts have a common translational motion, there is a rotational effect to be observed. The experimental arrangement is shown schematically in fig. 4. The semi-silvered plate divides the beam into two portions—one traversing the quadrilateral formed by the mirrors in the direction of rotation, the other in the opposite direction. An interference pattern is formed when the beams unite, just as in the Michelson-Morley experiment. The interference fringes are registered on a photographic plate attached to the apparatus.

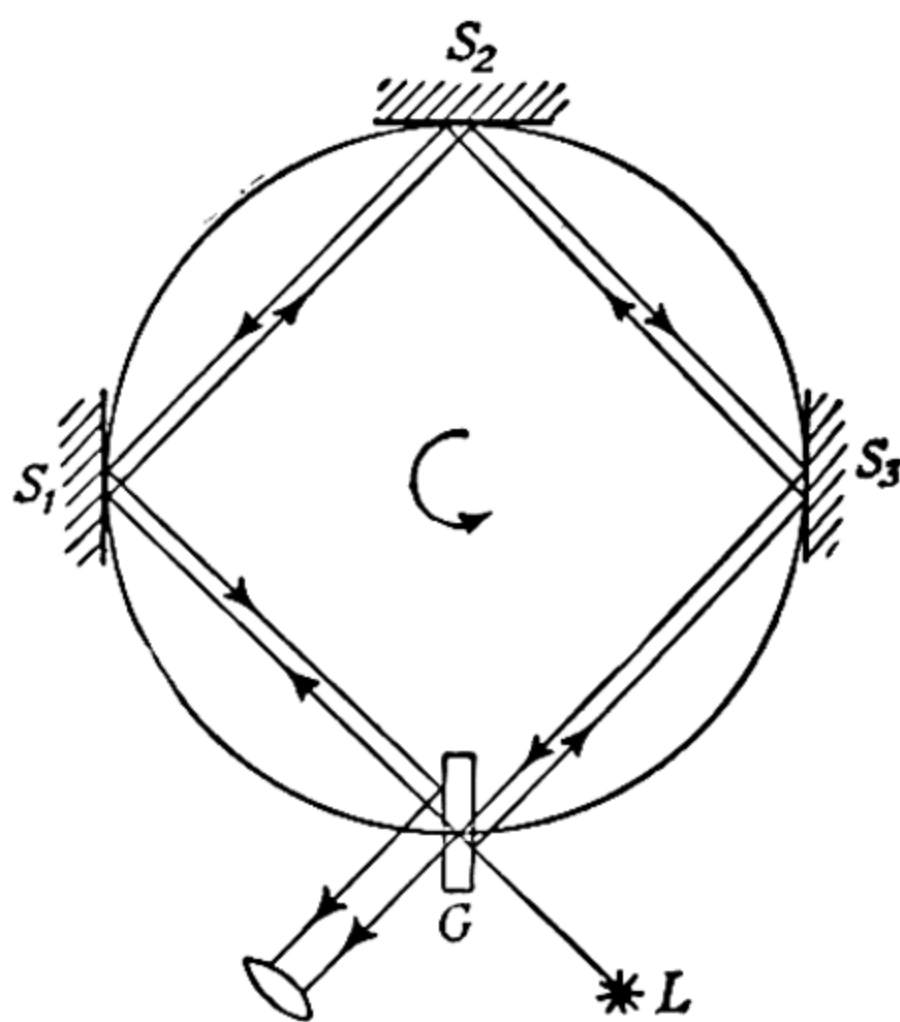


Fig. 4

As in the Michelson-Morley experiment, assume the classical compounding of velocities and compute the times required for the light to go in both directions and arrive again at G. Taking the axis of rotation as origin, suppose the light travels along a curve whose polar equation is  $r = r(\phi)$ . If, at a given instant, an element of the curve makes an angle  $\theta = \cos^{-1}(r d\phi/ds)$  with the direction of its translational motion arising from the turning of the

\* G. Sagnac, *Comptes rendus*, 157, p. 708 (1913); *Journ. de Physique*, 4, Ser. 5, p. 177 (1914).



whole apparatus, then the relative speed when going around in the same direction as the rotation will be

$$v_- = c - \omega r \cos \theta = c - r^2 \omega \frac{d\phi}{ds},$$

and in the opposite direction

$$v_+ = c + r^2 \omega \frac{d\phi}{ds}.$$

The difference of the two corresponding time intervals is

$$T_- - T_+ = \oint \frac{ds}{c - r^2 \omega \frac{d\phi}{ds}} - \oint \frac{ds}{c + r^2 \omega \frac{d\phi}{ds}}.$$

Since the second term in each denominator is small, this becomes

$$\Delta T = \oint \frac{2r^2 \omega d\phi}{c^2} = \frac{2\omega}{c^2} \oint r^2 d\phi = \frac{4\omega}{c^2} S,$$

where  $S$  is the area enclosed. Then, by p. 239, the displacement of the fringes amounts to

$$\Delta Z = \frac{4}{c\lambda} \omega S, \quad . \quad . \quad . \quad . \quad . \quad . \quad (19)$$

where  $\omega$  is the angular velocity of rotation,  $S$  the area of the surface enclosed by the light path and  $\lambda$  the wave-length of the light. This displacement is, in fact, observed. Experiments were performed using widely different values of both  $\omega$  and  $S$ . In one trial Sagnac used a rapidly rotating disc of about 1 m. diameter; in another trial the apparatus was taken on board ship; the angular motion was obtained by sailing over a curved course. In recent years Michelson and Gale \* used the rotation of the earth itself; in order to compensate for the small value of  $\omega$ , a light path several kilometres in length was used. Technically, this experiment differs from that of Sagnac in that the zero position of the fringes cannot be determined by stopping the rotation; hence it was necessary to have a second light path, enclosing a negligible area, so arranged as to give fringes serving as a reference mark. Further, unless the circuit is at one of the poles, the angular velocity vector is not normal to the plane of the circuit, and only the normal component is effective. The result of this experiment was in complete accord with the calculations. The fact that the stationary ether concept leads to correct results for rotating systems in optical experiments also is no more a proof of the existence of the ether than are the mechanical phenomena of the Coriolis and centrifugal forces (cf. p. 233). The Michelson-Gale experiment is the optical analogue of

\* A. A. Michelson and H. G. Gale, *Astrophys. Journ.*, 61, p. 140 (1925).

the Foucault pendulum. It is of importance from the fact that it makes all attempts to explain the negative result of the Michelson-Morley experiment by convection of the ether seem futile, for it seems absurd that the ether could be carried along by the earth completely in translation and not at all in rotation. As explained on p. 258, the above-mentioned mechanical experiments are explicable under the generalized theory of relativity without using the concept of an ether at rest, and the same is true of the Sagnac experiment.

#### 4.\* Relativistic-invariant Form of the Electromagnetic Equations.

The mathematical formulation of the theory of a stationary ether involves the application of formulæ (1) to (5) (p. 471), with the derivatives  $\partial/\partial t$  . . . for the moving bodies formed according to the Galilean transformation; i.e. we equate the total time rate of change of a quantity  $u$  to the sum of the pure time rate  $\partial u/\partial t$  for a fixed point and the spatial change  $\mathbf{v} \text{ grad } u$  due to displacement of this point, as in hydromechanics. But, as already indicated, this theory leads to conflicts with experience for effects of second order. Small as are the differences between the theoretical and experimental results, they nevertheless compel us to abandon the theory. Now it is very remarkable that, as will be shown below, the equations of the electron theory may be set down at once as equations connecting world vectors, which means that they are invariant under the Lorentz transformation and not under the Galilean. One might at first believe this to be a matter of course, since the Lorentz transformation was introduced in such a way that the wave equation remained invariant under it. It is to be noted, however, that while the wave equation is a consequence of the electromagnetic field equations, the electromagnetic equations are not, reciprocally, consequences of the wave equation, since the latter also occurs elsewhere, e.g. in the Theory of Elasticity. Again, the invariance of the electromagnetic equations under the Lorentz transformation—which is, as a matter of fact, identical with the Galilean to a first approximation—shows that the stationary ether theory also fails to indicate *first order* effects when all parts of an electromagnetic system share a common motion. This fact was stated in the preceding section, but without proof.

In order to obtain a simple form of equations (1) to (4) (p. 471), we restrict our considerations to weakly magnetic substances, so that the term  $\partial \mathbf{M}/\partial t$  in equation (2) may be omitted. In addition, let us combine the displacement current  $\partial \mathbf{P}/\partial t$  with the conduction current, i.e. let the terms  $(4\pi/c)\rho_i \mathbf{v}_i$  consist of the sum of a convection current of free charges and a part due to the displacement of the bound charges from their equilibrium positions. The term  $\partial \mathbf{P}/\partial t$  then appears to be absent, but actually  $\rho \mathbf{v}$  includes both the conduction current and the

contribution of the displacement. Writing  $l/ic$  in place of  $t$ , the coordinate forms of equations (1) and (3) become

$$\left. \begin{aligned} \frac{\partial H_z}{\partial y} - \frac{\partial H_y}{\partial z} - \frac{\partial iE_x}{\partial l} &= \frac{4\pi}{c} (\rho_+ v_{+x} - \rho_- v_{-x}) \\ - \frac{\partial H_z}{\partial x} + \frac{\partial H_x}{\partial z} - \frac{\partial iE_y}{\partial l} &= \frac{4\pi}{c} (\rho_+ v_{+y} - \rho_- v_{-y}) \\ + \frac{\partial H_y}{\partial x} - \frac{\partial H_x}{\partial y} - \frac{\partial iE_z}{\partial l} &= \frac{4\pi}{c} (\rho_+ v_{+z} - \rho_- v_{-z}) \\ + \frac{\partial iE_x}{\partial x} + \frac{\partial iE_y}{\partial y} + \frac{\partial iE_z}{\partial z} &= 4\pi i (\rho_+ - \rho_-) \end{aligned} \right\}, \quad (20)$$

while equations (2) and (4) are

$$\left. \begin{aligned} \frac{\partial iE_z}{\partial y} - \frac{\partial iE_y}{\partial z} - \frac{\partial H_x}{\partial l} &= 0 \\ \frac{\partial iE_x}{\partial z} - \frac{\partial iE_z}{\partial x} - \frac{\partial H_y}{\partial l} &= 0 \\ \frac{\partial iE_y}{\partial x} - \frac{\partial iE_x}{\partial y} - \frac{\partial H_z}{\partial l} &= 0 \\ \frac{\partial H_x}{\partial x} + \frac{\partial H_y}{\partial y} + \frac{\partial H_z}{\partial z} &= 0 \end{aligned} \right\} \dots \dots \dots (21)$$

We proceed to show that these are the component forms of two equations between world vectors, or tensors. First we investigate how the density of charge  $\rho$  changes under the Lorentz transformation. Start from the assumption that the charge  $e = \rho \Delta \tau$  is independent of the state of motion, and so does not change in this transformation. On the other hand, all volumes are reduced in the ratio  $\sqrt{1 - \beta^2} : 1$  by the Lorentz contraction; hence to a moving observer the charge density  $\rho$  appears to be increased in the ratio  $1/\sqrt{1 - \beta^2} : 1$ . Denoting by  $\rho_0$  the density of charge as measured by an observer moving with it, we have

$$\rho = \frac{\rho_0}{\sqrt{1 - \beta^2}} \quad \text{or} \quad \rho \sqrt{1 - \beta^2} = \rho_0 = \text{invariant.} \quad (22)$$

According to p. 256, the components of the four-velocity  $q^r$  are

$$\frac{v_x}{\sqrt{1 - \beta^2}}, \quad \frac{v_y}{\sqrt{1 - \beta^2}}, \quad \frac{v_z}{\sqrt{1 - \beta^2}}, \quad \frac{ic}{\sqrt{1 - \beta^2}}.$$

If we multiply each of these expressions by the invariant of (22) we obtain the components of a new world vector which we call the



*four-current*  $\mathbf{s}^r$ ; the positive charges yield the four-current  $\mathbf{s}^r_+$ —the negative, the four-current  $\mathbf{s}^r_-$ . Since the factor  $\sqrt{1 - \beta^2}$  cancels out in the multiplication, the components are the right members of equations (20), apart from a factor  $4\pi/c$ . Consequently, the left members must also be the components of a world vector. This is derived from a world tensor by an operation which is independent of the system of coordinates. We already applied such an operation in three dimensions in the Theory of Elasticity without introducing any special designation for it. If, in the three-dimensional case, say, we have a tensor in the form

$$\Psi = a_1 \cdot i + a_2 \cdot j + a_3 \cdot k,$$

and we use this tensor as a postfactor in multiplying with the nabla operator, we obtain

$$\begin{aligned} \nabla \Psi &= \nabla a_1 \cdot i + \nabla a_2 \cdot j + \nabla a_3 \cdot k \\ &= \text{div } a_1 \cdot i + \text{div } a_2 \cdot j + \text{div } a_3 \cdot k, \end{aligned}$$

i.e. a vector whose components are  $\text{div } a_1, \text{div } a_2, \text{div } a_3$ . We shall refer to this operation as the formation of the divergence vector of the tensor  $\Psi$ , and shall designate the operation by

$$\nabla \Psi = \text{div } \Psi. \quad . \quad . \quad . \quad . \quad . \quad . \quad (23)$$

The extension to the four-dimensional case involves nothing new. If we call the world tensor whose components are

0	$H_z$	$-H_y$	$-iE_x$
$-H_z$	0	$H_x$	$-iE_y$
$H_y$	$-H_x$	0	$-iE_z$
$iE_x$	$iE_y$	$iE_z$	0

the electromagnetic field tensor  $\Phi^r$ , we see that the components of the world vector of equation (20) are obtained from this world tensor  $\Phi$  precisely by the operation given above. Hence equation (20) may be written

$$\text{div}^r \Phi^r = \frac{4\pi}{c} (\mathbf{s}^r_+ + \mathbf{s}^r_-). \quad . \quad . \quad . \quad . \quad (24)$$

But according to equation (65) (p. 251), the system (21) above states that the antisymmetric world tensor  $\Phi^r$  is derived from a world vector potential  $\mathbf{V}^r$  by taking the curl in four dimensions:

$$\Phi^r = \text{curl}^r \mathbf{V}^r. \quad . \quad . \quad . \quad . \quad . \quad (25)$$

The formulation of the laws of electromagnetism given by (24) and (25) seems, at first, somewhat colourless; nevertheless, compared with the usual three-dimensional vector equations, it has the great advantage that it combines the electric and magnetic fields so as to produce

a single unified field specified by the tensor  $\Phi^r$ , the resolution into electric and magnetic components depending upon the state of motion of the observer. A field which appears to an observer at rest to be purely magnetic may, for a moving observer, appear to have an electrical component. Moreover, the change from one system to another is reduced to the purely formal operation of transforming four-dimensional tensor components. For example, in the experiment of W. Wien (p. 474), the world tensor  $\Phi^r$  has the following components for the system in which the apparatus is at rest:

$$\begin{array}{cccc} 0 & H & 0 & 0 \\ -H & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0. \end{array}$$

On the other hand, consider a canal ray particle moving in the direction of the  $x$ -axis with velocity  $v$ . Since

$$\begin{aligned} l' &= l \cos \phi + x \sin \phi, \\ x' &= -l \sin \phi + x \cos \phi, \\ \tan \phi &= -i\beta, \end{aligned}$$

the rule for transforming of tensor components (pp. 34, 249) yields

$$\begin{array}{cccc} 0 & H \cos \phi & 0 & 0 \\ -H \cos \phi & 0 & 0 & -H \sin \phi \\ 0 & 0 & 0 & 0 \\ 0 & H \sin \phi & 0 & 0 \end{array}$$

We see that an electrical component has made its appearance, viz.:

$$iE_v' = H \sin \phi = -\frac{Hi\beta}{\sqrt{1-\beta^2}} \quad \text{or} \quad E_v' = -\frac{Hv/c}{\sqrt{1-\beta^2}},$$

which is identical with that derived from equation (4), p. 471, as far as quantities of second order. Likewise, the magnetic component remains unchanged as far as second order quantities.

*Ex. 117.* Derive the Fresnel coefficient from the relativistic velocity addition theorem.





## PART V

### THE THEORY OF HEAT. PHENOMENOLOGICAL PART

The Theory of Heat, like the Theory of Electricity, may be developed from two different points of view. The phenomenological method employs only concepts like temperature, quantity of heat, &c., which are taken from the macroscopic world of observations and which can be measured directly. The laws thus obtained have the advantage of being free from hypothetical assumptions. On the other hand, in the Theory of Heat, e.g. in connexion with the Law of Entropy, we feel the need of a deeper "explanation". Such an interpretation, deeper because more vivid, is furnished by the atomistic, statistical view. Here again, as in the electrical case, we find that although we develop the theory separately from the two sides, it will often prove advantageous, while we follow one line of approach, not to shut our eyes completely to the other.



## CHAPTER XXVIII

### THEORY OF THE CONDUCTION OF HEAT

#### 1. Definition of Temperature, Quantity of Heat, Thermal Capacity and Specific Heat.

The sensory organs of the skin permit us, within certain limits, to perceive temperature. In particular, we can decide readily which of two objects is the warmer. However, this qualitative measure of temperature is not sufficient for physical measurements and we must seek a phenomenon which is causally connected in some way with the temperature of a body. The increase in volume of a substance, e.g. mercury, with rising temperature is a phenomenon of this kind. For convenience of observation of the expansion, the mercury is contained in a bulb and allowed to expand into a fine capillary tube attached to it. A thermometer of this kind is calibrated by immersing the bulb in melting ice and marking the position of the mercury filament, then immersing the bulb in the steam rising from water boiling at standard atmospheric pressure and again marking the position of the thread of mercury. The interval is then divided into one hundred equal parts. We have thus obtained a provisional scale of temperature. If, now, we repeat this process with another thermometric substance, we find that accurate observation reveals a difficulty; viz. since the volume is not strictly a linear function of temperature, two thermometers filled with different substances and calibrated as above will not agree exactly at intermediate points when immersed in the same substance. We might of course avoid this difficulty by the arbitrary choice of a standard thermometric substance, say mercury. There is, however, a better plan: we obtain almost perfect agreement by using gases which are as far as possible from their boiling-points. For practical reasons a gas thermometer is usually arranged so that we measure not the expansion but rather the increase in pressure occurring when the volume is held constant. Since individual differences arise when such gases are in the neighbourhood of their boiling temperatures, we choose as a thermometric substance that gas which, under ordinary conditions, is farthest from its boiling-point, viz. helium. Thus *accurate thermometric measurements are based upon a temperature scale determined by the increase in pressure of a con-*



*fixed volume of helium.* Moreover, the gas thermometer defines a natural zero point for our scale—the so-called absolute zero—which is at  $-273^{\circ}$  on the Centigrade scale defined above. The determination of this zero point will be discussed later. The temperature scale defined on this basis departs but little from one defined on purely theoretical grounds, the thermodynamic scale, which will be explained later. We shall denote temperatures on the Centigrade scale by  $t$ , those on the absolute scale by  $T$ . There will be no danger of confusing  $t$  with the time, since pure thermodynamics deals only with equilibrium. In the Theory of Heat Conduction, however, we shall denote the temperature by  $\theta$ . Since, in the latter case, we are concerned only with temperature *differences*, it is immaterial whether we understand this to be Centigrade or absolute temperature.

In addition to temperature there is a second important concept—that of quantity of heat—which is given by ordinary experience. If we drop a 1 gm. mass of iron at  $100^{\circ}$  C. into 10 gm. of water at  $10^{\circ}$  C. and, in a second experiment, drop a 10 gm. piece of iron at  $100^{\circ}$  C. into the same amount of water at  $10^{\circ}$  C. as before, we observe, after the systems have come to equilibrium, that the rise in temperature of the water is about nine times as great in the second case as in the first. This leads to the idea that a body possesses a given heat content which is equal to the product of the temperature by a quantity proportional to the mass. This latter quantity is called the heat capacity. It is to be noticed that we deal only with *differences* in quantities of heat, so that the absolute value is undetermined. The process of coming to equilibrium in the above example may then be interpreted as the transfer of a given quantity of heat from one body to another. Without forming special conceptions of the nature of heat, we postulate a *conservation of quantity of heat* for such processes. As a unit of quantity of heat, we choose the amount of heat necessary to raise the temperature of one gramme of water from  $14.5^{\circ}$  C. to  $15.5^{\circ}$  C., and call this a small calorie (sometimes “gramme calorie”). There is a larger unit, the large calorie (or kilogram calorie) which is one thousand times as great.

Experimenting further, we find that 1 gm. of aluminium at  $100^{\circ}$  C., when placed in 10 gm. of water at  $10^{\circ}$  C. imparts to it a greater rise in temperature than does 1 gm. of lead at  $100^{\circ}$  C. Hence we set the heat capacity equal to the product of the mass by a constant of the material. This constant is taken to be unity for water at  $15^{\circ}$  C. The quantity of heat in calories required to raise the temperature of 1 gramme of a substance  $1^{\circ}$  C. is called the *specific heat* of the substance, preferably called the *specific heat per gramme*. There is another quantity which is important—the specific heat per mol, i.e. the thermal capacity of one mol of a substance. The advantage in using this *molecular heat* is that in comparing the molecular heats of various substances, the same

numbers of molecules are involved. We shall denote the thermal capacity by  $C$ , the specific heat per gramme by  $c$ , and the molecular heat by  $c$ . In order to determine the equilibrium temperatures in the above examples we need only equate the quantities of heat involved, before and after mixing. In the example of 1 gm. of iron at  $100^\circ$  C. placed in 10 gm. of water at  $10^\circ$  C., the end temperature  $x$  is obtained from the relationship

$$1 \cdot c_g \cdot 100 + 10 \cdot 10 = (c_g + 10)x,$$

where  $c_g$  is the specific heat per gramme of the iron.\*

In spite of this simple applicability of the concept of heat content, it must be mentioned that these simple relationships exist only when the bodies under consideration do not undergo other changes. In general, the heat capacity is by no means a function of the temperature alone (cf. p. 502).

## 2. The Differential Equation of Heat Conduction; Initial and Boundary Conditions.

Common experience shows that any temperature difference between parts of a body reduces to zero after a time, so that heat must flow from the places of higher temperature to those of lower temperature. It is natural to assume that the density of the thermal current, i.e. the quantity of heat passing per second through unit area of a surface normal to the direction of flow, is proportional to the negative of the temperature *gradient*. This assumption is confirmed by experiment. The factor of proportionality  $k$  is called the *thermal conductivity* of the substance. If now we imagine a certain volume of a body isolated, the rate of decrease of the heat content is

$$-\frac{\partial W}{\partial t} = -\frac{\partial}{\partial t} \int \rho \theta c_g d\tau, \quad . . . . . (1)$$

where  $\theta$  is the temperature and  $\rho$  the density. This decrease is due to conduction, the heat flowing away through the bounding surface. The flow is given by

$$-\frac{\partial W}{\partial t} = -k \oint \text{grad } \theta dS. \quad . . . . . (2)$$

Equating the two, we have

$$\frac{\partial}{\partial t} \int \rho \theta c_g d\tau = k \oint \text{grad } \theta dS = k \int \text{div grad } \theta d\tau. \quad . (3)$$

\* This is equivalent to equating the heat given up by the hot body (or bodies) to that absorbed by the cold, in which case the heat equation is

$$1 \cdot c_g (100 - x) = 10(x - 10),$$

which is, of course, the same as the equation given above.



Since this must hold for every element of volume, we have the equation of heat conduction

$$\frac{\partial \theta}{\partial t} = \frac{k}{\rho c_0} \Delta \theta = m \Delta \theta. \quad . \quad . \quad . \quad . \quad . \quad (4)$$

The constant  $m = k/\rho c_0$  was named the *diffusivity* by Kelvin.\*

As in the case of all partial differential equations, the solution of the problem is not defined unless initial and boundary conditions are given. Thus there must be given a definite distribution of temperature at time  $t = 0$ , and also information as to what is happening at the boundaries of the body. The surface of the body, for example, may be completely insulated. In general this will not be the case; as a rule, heat will pass outwards through the surface. It is found that the surface density of the flow of heat normally outwards through the surface is approximately proportional to the difference in temperature between the surface of the body and the surrounding medium. The factor of proportionality is called the *surface conductivity*, or *emissivity*.

### 3. A Simple Example of the Integration of the Equation of Heat Conduction: Penetration of the Daily and Yearly Temperature Variations into the Interior of the Earth.

Take the surface of the earth to be plane and assume the surface to undergo periodic temperature variations as a result of the fluctuating radiation to which it is exposed. The boundary conditions here are especially simple, for there is no heat flow through the lateral surface of a cylinder whose generators are vertical. This is because the isothermal surfaces are planes parallel to the earth's surface, as may be seen intuitively; hence the entire flow is normal to the surface of the earth. Taking the  $x$ -axis along the inward normal to this surface, the differential equation becomes

$$\frac{\partial \theta}{\partial t} = m \frac{\partial^2 \theta}{\partial x^2}. \quad . \quad . \quad . \quad . \quad . \quad (4')$$

Since  $\theta$  at the surface is a periodic function of the time, we take

$$\theta = f(x) e^{i\omega t}. \quad . \quad . \quad . \quad . \quad . \quad (5)$$

Substituting this value in (4') we obtain the following differential equation for  $f(x)$ :

$$m \frac{d^2 f}{dx^2} - i\omega f = 0, \quad . \quad . \quad . \quad . \quad . \quad (6)$$

\* Clerk-Maxwell called this quantity the *thermometric conductivity*, but Kelvin's terminology is more widely used.



which we seek to solve by setting  $f = ae^{\lambda x}$ . The "characteristic equation" becomes

$$m\lambda^2 - i\omega = 0, \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

whence

$$\lambda = \pm \sqrt{\frac{i\omega}{m}} = \pm \sqrt{\frac{\omega}{2m}} (1 + i). \quad . \quad . \quad . \quad (8)$$

The negative sign is the one to be selected, for if we limit our inquiry to the penetration of the surface variations, there must be a *decrease* toward the interior. Then

$$\theta = ae^{-\sqrt{\omega/2m}x} e^{i\omega[t-(x/\sqrt{2m\omega})]}. \quad (9)$$

This is the equation of a damped wave propagated with velocity  $v = \sqrt{2m\omega}$ . This shows that the velocity is greater for the diurnal variation than for the annual, but at the same time the damping is greater for the former.

It is also easy to fulfil the initial time conditions for  $x = 0$ . Since the real as well as the imaginary part of (9) represents a solution, we may obtain the solution for any values of  $\theta_0$  and  $(\partial\theta/\partial x)_0$  from the general solution containing two constants of integration:

$$\theta = Ae^{-\sqrt{\omega/2m}x} \cos \omega \left( t - \frac{x}{\sqrt{2m\omega}} \right) + Be^{-\sqrt{\omega/2m}x} \sin \omega \left( t - \frac{x}{\sqrt{2m\omega}} \right). \quad \dots (10)$$

If, for example, we begin counting time from a moment when  $\theta$  at  $x = 0$  is a maximum, we have the solution

$$\theta = \theta_{\max} e^{-\sqrt{\omega/2m} x} \cos \omega \left( t - \frac{x}{\sqrt{2m\omega}} \right), \quad . \quad (10')$$

where  $\theta$  and  $\theta_{\max}$  are, of course, the *differences* reckoned from the average temperature.

## CHAPTER XXIX

### THE EQUATION OF STATE OF THERMODYNAMIC SYSTEMS

#### 1. Definition of the Thermodynamic Variables and the Relationships between them.

In thermodynamics, the treatment of which we begin here, the word "*system*" is used as the most general and neutral expression for an aggregation of bodies considered as a whole and assumed to be isolated from their surroundings. A system may consist of several distinct physical parts called *phases*. For example, if we have a saturated NaCl solution containing an undissolved crystal of NaCl, the whole being covered by a bell jar, then we have a system of three phases: a solid phase (the crystal), a liquid phase (the solution) and a gaseous phase consisting, in this instance, of a mixture of air and water vapour. Since all gases are miscible, a system in equilibrium can have but *one* gaseous phase, but may have several solid or liquid phases. For example, we can introduce an additional solid phase by placing a crystal of quartz in the solution. The thermodynamic behaviour of the system is dependent upon the number of "independent components" or "substances". We understand this to mean the number of freely variable chemical compounds or elements contained in the system. To obtain this number, determine first the number of elements present (combined or free) and subtract from this the number of elements defined chemically in terms of the remaining (independent) elements. For example, water and ice constitute a two-phase system of *one* component, since the quantity of oxygen present is determined by the quantity of hydrogen. On the other hand, a solution of NaCl in water is a single-phase system of two components. The state of a system is determined, in the thermodynamic sense, by giving the mass and chemical constitution of each phase as well as the pressure, volume and temperature. These quantities, however, are not independent of one another, but are connected by a relationship called an *equation of state*.

We limit our present considerations to a single-phase system whose chemical constitution is invariable, e.g. a system consisting of 10 gm. of liquid benzene. Since the mass and chemical constitution are fixed, there remain only the three quantities temperature  $t$  (or  $T$ ), pressure

$p$  and volume  $V$  as true thermodynamic variables. Temperature and methods of measuring it have already been discussed in Chap. XXVIII (p. 487). The volume is measured in cubic centimetres and the pressure in dynes per square centimetre. In some instances the pressure is measured in atmospheres, this being the pressure capable of supporting a 76 cm. column of mercury at  $0^{\circ}$  C.

$$1 \text{ Atm.} = 76 \times 13.6 \times 981 = 1.013 \times 10^6 \text{ dynes/cm.}^2 = 1.013 \times 10^5 \text{ newton/m.}^2$$

Now there is generally an equation of state

$$F(p, t, V) = 0, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

according to which the condition of a simple system is determined by fixing the values of two of the variables. An equation of state which is valid over large ranges can be given only for gases. For liquids and solids these equations are more like interpolation formulæ holding over a limited range. But, as we shall see presently, important conclusions may be drawn from the mere fact that an equation of state exists.

We now consider those changes of state in which each variable in turn is held fixed. We thus arrive at definitions for three coefficients, viz. the following:

The *isobaric volume coefficient of expansion*  $\alpha$  is defined as the rate of change of volume per degree divided by the volume at 0° C., the pressure being held constant, i.e.

$$\alpha = \frac{1}{V_0} \lim_{\Delta t \rightarrow 0} \frac{V(p, t + \Delta t) - V(p, t)}{\Delta t} = \frac{1}{V_0} \left( \frac{\partial V}{\partial t} \right)_p. \quad (2)$$

The subscript attached to the partial derivative indicates the quantity which is held constant.

The *isometric* (or *isochoric*) *pressure coefficient*  $\beta$  is defined as the rate of pressure change per degree divided by the pressure at 0° C., the volume remaining constant:

$$\beta = \frac{1}{p_0} \lim_{\Delta t \rightarrow 0} \frac{p(V, t + \Delta t) - p(V, t)}{\Delta t} = \frac{1}{p_0} \left( \frac{\partial p}{\partial t} \right)_V. \quad (3)$$

Finally, the *isothermal compressibility*  $\kappa$  is defined as the rate of change in volume per unit pressure (or per atmosphere) divided by the existing volume, the temperature being held constant:

$$\kappa = -\frac{1}{\bar{V}} \lim_{\Delta p} \frac{V(t, p + \Delta p) - V(t, p)}{\Delta p} = -\frac{1}{\bar{V}} \left( \frac{\partial V}{\partial p} \right)_t. \quad (4)$$

Since the volume decreases with increasing pressure we insert the negative sign to make  $\kappa$  positive. The reciprocal of  $\kappa$  is called the *isothermal volume elasticity*  $E_t$ . It is to be noted that, for reasons of convenience,



the coefficients are so defined that  $\alpha$  and  $\beta$  are referred to the volume or pressure at  $0^\circ \text{C.}$ , while  $\kappa$  is referred to the existing volume. The advantage of this convention becomes apparent in the case of gases. For liquids and solids it is practically immaterial whether the change is referred to the volume at  $0^\circ \text{C.}$  or the volume at the existing temperature. Again, it is immaterial whether we write  $t$  or  $T$  in defining  $\alpha$  and  $\beta$ , since temperature *differences* only are concerned. The existence of an equation of state means that the three coefficients are not independent. If, for example, we imagine the equation solved for  $p$ , then

$$p = p(V, t)$$

and 
$$dp = \left( \frac{\partial p}{\partial V} \right)_t dV + \left( \frac{\partial p}{\partial t} \right)_V dt. \quad . . . . (5)$$

If  $p$  is held constant we have, after division by  $dt$ ,

$$0 = \left( \frac{\partial p}{\partial V} \right)_t \left( \frac{\partial V}{\partial t} \right)_p + \left( \frac{\partial p}{\partial t} \right)_V.$$

Inserting the coefficients defined above in place of the derivatives,

$$\beta = \frac{\alpha}{\kappa} \cdot \frac{V_0}{p_0 V}. \quad . . . . (6)$$

*Ex. 118.* A mercury thermometer whose capillary is just filled at  $45^\circ \text{C.}$  is warmed to  $50^\circ$ . What is the pressure in the capillary if the coefficient of expansion  $\alpha = 18 \times 10^{-6} \text{ deg.}^{-1}$  and the compressibility  $\kappa = 39 \times 10^{-7} \text{ Atm.}^{-1}$ ?

## 2. The Equation of State of an Ideal Gas.

The following empirical relationships have been found to hold for gases far removed from their point of liquefaction:

1. *The Law of Boyle and Mariotte.*—At constant temperature the volume of a gas is inversely proportional to the pressure, i.e. the product of pressure by volume depends only upon the temperature:

$$pV(p, t) = p_0 V(p_0, t), \quad . . . . (7)$$

where  $p_0$  denotes some standard pressure, say 1 atm.

2. *The Law of Gay-Lussac.*—The isobaric volume coefficient has the same value  $\alpha = 1/273.1$ , independent of the temperature, for all gases. Thus by the defining equation (2) (p. 493),

$$V(p, t) = V(p, 0) (1 + \alpha t), \quad . . . . (8)$$

and in particular

$$V(p_0, t) = V(p_0, 0) (1 + \alpha t). \quad . . . . (8')$$

Combining (7) with (8') we obtain

$$pV(p, t) = p_0 (1 + \alpha t) V(p_0, 0). \quad . . . . (9)$$

We now change the zero of the temperature scale from the temperature of melting ice to the temperature  $-1/\alpha = -273.1^\circ \text{C.}$ , and take

$$t = T - 273.1 = T - \frac{1}{\alpha}. \quad . \quad . \quad . \quad . \quad (10)$$

Then we have

$$pV = \frac{p_0 V(p_0, 273.1) T}{273.1} = \frac{p_0 V_0}{273.1} T = AT. \quad . \quad . \quad (11)$$

We have already referred to this scale of temperature as the absolute scale. The constant  $A$ , which is equal to  $1/273.1$  of the product of the pressure and volume of the gas at  $0^\circ \text{C.}$ , varies according to the mass and nature of the gas under consideration. However, we obtain a universal gas equation by using another fact of experience as follows.

3. *The Law of Avogadro.*—The densities of two gases which are at the same pressure and temperature are proportional to the molecular weights. If  $M$  is the mass of the gas, the density is  $M/V$  and so

$$\frac{M_1}{V_1} : \frac{M_2}{V_2} = m_1 : m_2 \quad . \quad . \quad . \quad . \quad (12)$$

where  $m_1, m_2$  are the molecular weights. Equation (12) is equivalent to

$$\frac{m_1 V_1}{M_1} = \frac{m_2 V_2}{M_2}. \quad . \quad . \quad . \quad . \quad (12')$$

The quantity

$$\frac{mV}{M} = v$$

is the volume occupied by one mol. This, as well as all symbols referred to one mol (gramme molecule), will be denoted by small letters.

If we write the gas equation for one mol ( $V = v$ ), the constant on the right becomes the same for all gases, since by Avogadro's Law, one mol of *any* gas occupies a given volume  $v_0$  at a given temperature and pressure. We then have

$$pv = \frac{p_0 v_0}{273.1} T = RT. \quad . \quad . \quad . \quad . \quad (13)$$

The numerical value of this so-called universal gas constant is

$$R = 8.314 \times 10^7 \text{ erg/deg. mol} = 8310 \text{ int. joule/deg. kg.-mol.}$$

Since the mass of one mol is  $m$  grammes, where  $m$  is the molecular weight, multiplication by  $M/m$  yields the following result for any mass  $M$  of gas

$$pV = \frac{M}{m} RT = nRT, \quad . \quad . \quad . \quad . \quad (13')$$

where  $n$  is the number of mols of gas present.

The product  $pv$  vanishes for  $T = 0$ . This means that an "ideal" gas which obeys (13) would have zero volume for a finite pressure, or would no longer exert pressure, the volume still being finite. The former alternative seems impossible to visualize. Now the atomistic view reveals at once the nature of the simplification inherent in the definition of an ideal gas which leads to such paradoxical results. According to the kinetic theory of gases, the pressure exerted by a gas on the walls of its container is due to the innumerable impacts of the rapidly-moving molecules. The molecules themselves may be assumed to be particles having no extension, so that the entire volume  $v$  is available to every molecule. The result is that at absolute zero, where the molecules no longer are in motion, the volume of the gas is zero. In reality we know, however, that the colliding molecules behave like elastic spheres whose diameter is of the order of  $10^{-8}$  cm., so that they occupy a small, but nevertheless finite, volume. Further, in deriving the gas equation from the kinetic theory, it is assumed that the molecules exert no forces on each other. We know, however, that large intermolecular forces of attraction exist in liquids and solids (cohesive forces) and it is thus probable that such forces exist in the gaseous state also, but that they are much feebler, corresponding to the greater average distance between molecules.

We may now derive the equation of state of a mixture of ideal gases. Let the masses  $M_1, M_2, M_3, \dots$  of gases having molecular weights  $m_1, m_2, \dots$  be contained in a volume  $V$ . In order to obtain the equation of state for this mixture we need the following law.

4. *Dalton's Law of Partial Pressures*.—The pressure exerted by a mixture of gases is equal to the sum of the partial pressures which each gas itself would exert if it filled the entire volume. From the kinetic theory standpoint this law is a matter of course, since the total effect of the collisions with the walls must be equal to the sum of the effects due to the separate kinds of molecules. Thus for the first component gas,

$$p_1 V = \frac{M_1}{m_1} RT = n_1 RT,$$

and similarly,

$$p_2 V = \frac{M_2}{m_2} RT = n_2 RT.$$

Adding all such equations and applying Dalton's Law, we obtain

$$pV = \left( \frac{M_1}{m_1} + \frac{M_2}{m_2} + \dots \right) RT = (n_1 + n_2 + n_3 + \dots) RT = nRT \quad (14)$$

and

$$p_i = \frac{n_i}{n_1 + n_2 + n_3 + \dots} p. \quad \dots \quad (14a)$$

Using this relationship we can define the apparent molecular weight



of the mixture as the molecular weight  $\bar{m}$  of a gas which, for the same mass  $M$ , would satisfy the same equation of state, viz. the equation

$$pV = \frac{M}{\bar{m}} RT = \frac{M_1 + M_2 + M_3 + \dots}{\bar{m}} RT. \quad (15)$$

Comparison of (14) and (15) gives the apparent molecular weight of the mixture as

$$\bar{m} = \frac{M_1 + M_2 + M_3 + \dots}{\frac{M_1}{m_1} + \frac{M_2}{m_2} + \frac{M_3}{m_3} + \dots}. \quad (16)$$

### 3. The Equation of State of a Real Gas.

The atomistic view shows us how to improve the equation of state of a gas: it is necessary only to take account of the quantities which were neglected in arriving at the equation of an ideal gas. Firstly, we must subtract from the macroscopic volume  $v$  a quantity  $b$  which allows for the space occupied by the molecules. This quantity  $b$  is four times the volume actually occupied by the molecules, which are assumed to be spherical. The reason for this is that in a collision the centres of two molecules approach no nearer than a diameter, so that for impact we may imagine half the molecules to be mere points, while the remainder occupy eight times their actual volume. Hence, for the average, we use four times the true volume of the molecules.

The cohesive forces operating on any given molecule in the interior of the gas cancel out, since they are essentially equal in all directions; not so, however, for a molecule near a wall of the vessel. It will be drawn back somewhat into the interior of the gas, if we assume no attraction between the gas molecules and the wall (adhesion). This has the effect of a small inward pressure in addition to the observed manometric pressure  $p$ . This added pressure must be inversely proportional to the square of the volume, for the force acting on a given colliding molecule is proportional to the number of molecules in the interior, which, in turn, is proportional to the density of the gas. But the pressure increase is also proportional to the number of molecules colliding with the walls, and hence again proportional to the density. Thus, in place of Boyle's Law we write the following equation for one mol of gas:

$$\left(p + \frac{a}{v^2}\right)(v - b) = \left(p_0 + \frac{a}{v_0^2}\right)(v_0 - b). \quad (17)$$

The product is again set proportional to the absolute temperature, but the constant  $R$  is to be replaced by gas constants  $R'$  which differ somewhat for various gases. The resulting equation of state is named

after van der Waals, to whom most of the above considerations are due:

$$\left(p + \frac{a}{v^2}\right)(v - b) = R'T. \quad . \quad . \quad . \quad (18)$$

In order to obtain a view of the relationships embodied in this formula we draw the lines  $T = \text{const.}$  in the  $vp$ -plane (fig. 1). These lines are called *isothermal lines*, or simply *isotherms*. The isotherms of an ideal gas are the rectangular hyperbolas  $pv = \text{const.}$  The greater the value of  $T$ , the less important are the correction terms in van der Waals' equation. This means that the isotherms for higher values of  $T$  are not sensibly different from the hyperbolas of an ideal gas. For small

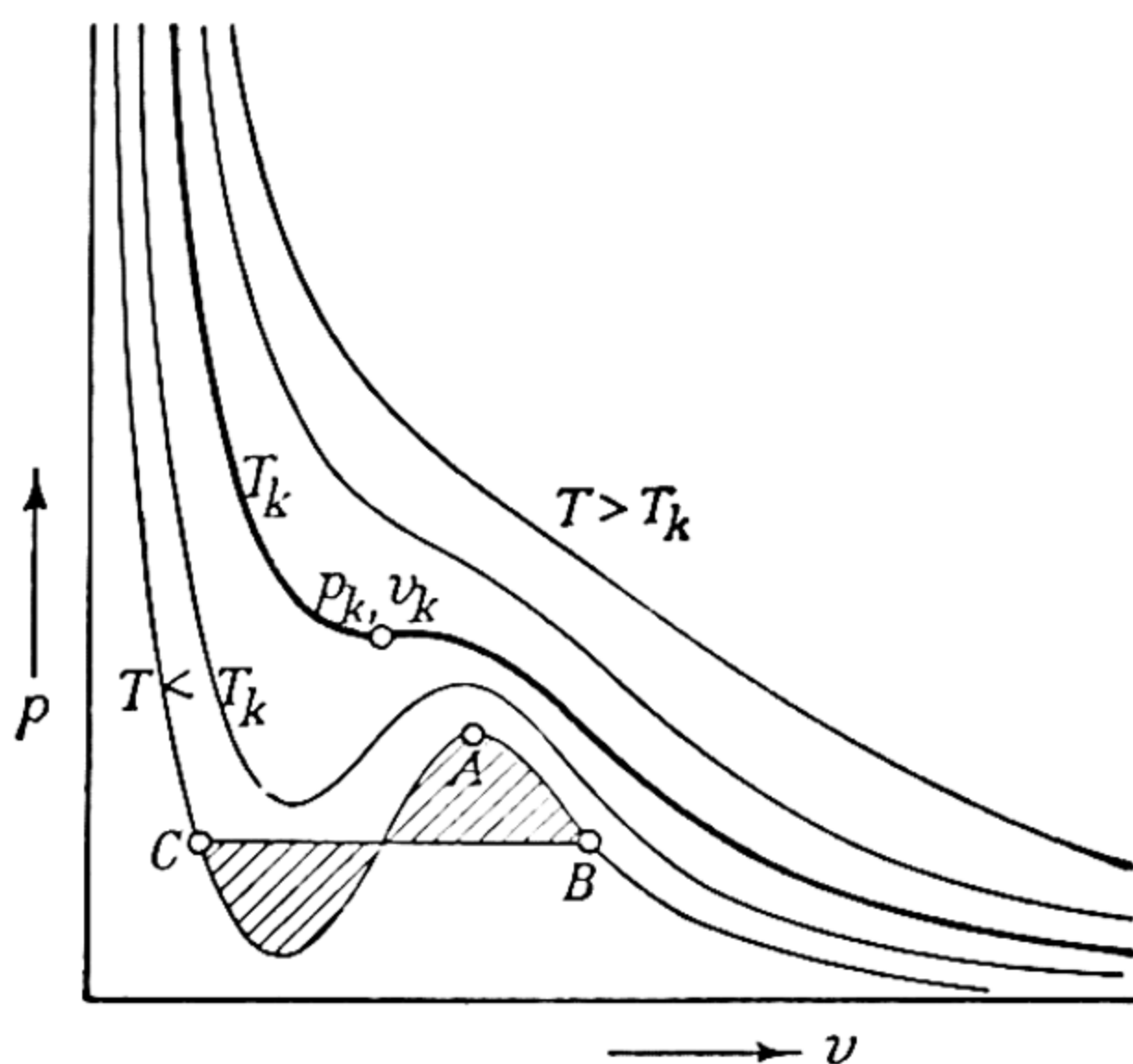


Fig. 1

values of  $T$ , on the other hand, the isotherms are of a different type, each one having a maximum and a minimum of  $p$ . The transition between the two types is a curve having a point of inflection with a horizontal tangent at which both extrema coincide. If, now, the point representing the state of a gas moves inward from large values of  $v$  along a low isotherm, the pressure increases as the volume is decreased. This is necessary in order to reach equilibrium, since we may imagine the volume to be decreased by loading a piston which confines the gas in a cylinder. After we reach the maximum of the curve, however, the pressure begins to fall as the volume is further decreased. There is an unstable region from  $A$  onward, since the internal pressure no longer increases if we seek to decrease the volume by loading the piston. Actually, this portion of the curve is not traversed at all; part of the gas begins to liquefy at  $B$ , and the pressure remains constant as the volume is further decreased. In the meantime, more and more gas is liquefied, until it is all in the liquid state at  $C$ ,

and the remaining branch of the curve gives the connexion between volume and pressure of the liquid. We shall find later that the horizontal line is to be so drawn that the two shaded areas are equal (cf. *Ex.* 124, p. 524).

Curves having no extrema can have no regions of this kind; thus is explained the fact that above a certain *critical temperature* it is impossible to liquefy a gas, no matter how great the applied pressure. The corresponding curve is the one having the point of inflection. The co-ordinates  $T_k$ ,  $p_k$ ,  $v_k$  of this point are called the critical temperature, critical pressure and critical volume. The following equations hold for these quantities:

$$p_k = \frac{R'T_k}{v_k - b} - \frac{a}{v_k^2}, \quad \dots \dots \dots (19)$$

and since the tangent is horizontal,

$$\left(\frac{\partial p}{\partial v}\right)_k = 0 = -\frac{R'T_k}{(v_k - b)^2} + \frac{2a}{v_k^3}. \quad \dots \dots (20)$$

Since there is a point of inflection,

$$\left(\frac{\partial^2 p}{\partial v^2}\right)_k = 0 = \frac{2R'T_k}{(v_k - b)^3} - \frac{6a}{v_k^4}. \quad \dots \dots (21)$$

These equations are readily solved for the critical values; for example, division of (20) by (21) yields

$$v_k = 3b, \quad \dots \dots \dots (22)$$

whence, from (20),

$$T_k = \frac{8a}{27R'b}, \quad \dots \dots \dots (23)$$

and from (19),

$$p_k = \frac{a}{27b^2}. \quad \dots \dots \dots (24)$$

Combining the last three equations gives the relationship

$$\frac{p_k v_k}{R'T_k} = \frac{3}{8} \quad \dots \dots \dots (25)$$

for the critical point. For an ideal gas, the expression  $pv/RT$  is equal to unity.

It is also possible to express the three constants of the van der Waals equation in terms of the critical constants:

$$\left. \begin{aligned} b &= \frac{v_k}{3}, & a &= 3p_k v_k^2 \\ R' &= \frac{8}{3} \frac{p_k v_k}{T_k} \end{aligned} \right\} \dots \dots \dots (26)$$



The equation of state then becomes

$$\left(p + \frac{3p_k v_k^2}{v^2}\right) \left(v - \frac{v_k}{3}\right) = \frac{8}{3} \frac{p_k v_k T}{T_k}. \quad \cdot \cdot \cdot \quad (27)$$

If we now introduce the following "reduced variables":

$$\Pi = \frac{p}{p_k}, \quad \omega = \frac{v}{v_k}, \quad \tau = \frac{T}{T_k}, \quad \cdot \cdot \cdot \quad (28)$$

we obtain an equation free of the constants of individual gases:

$$\left(\Pi + \frac{3}{\omega^2}\right) (3\omega - 1) = 8\tau. \quad \cdot \cdot \cdot \quad (29)$$

The existence of a reduced equation of state depends solely upon our being able to represent the facts by a formula containing three constants.\* This is possible only to a limited extent, since many substances undergo chemical changes as the density increases (e.g. association). Such changes are not accounted for by the van der Waals formula or by any other three-constant equation, since they are entirely of an individual nature. The existence of a general equation of the type (29) is known as the Theorem of Corresponding States. While the above considerations show that this theorem may not be valid in some cases, it may usually be employed advantageously in dealing with chemically similar substances.

*Ex. 119.* Derive expressions for the compressibility (or for the volume modulus of elasticity) of an ideal gas and of a van der Waals gas.

\* A reduced equation of state is still obtained if an individual gas constant is used in place of the universal constant  $R$ . The former, too, is eliminated in obtaining the reduced equation.

## CHAPTER XXX

### THE FIRST LAW OF THERMODYNAMICS: THE CONSERVATION OF ENERGY

#### 1. Formulation of the First Law.

The First Law makes a statement concerning the nature of heat. It asserts that *heat is a form of energy*, and that *in a closed system the total store of energy, i.e. the sum of the heat energy and mechanical or electrical energy, is constant*. This is the *Principle of the Conservation of Energy*, formulated by Robert Mayer in 1842. Thus, in a closed system, one form of energy may be transformed into another, but the sum must remain constant. In the atomistic treatment of the theory of heat, we shall see that heat is a manifestation of the kinetic energy of the atoms and molecules of the substance, the difference between this and the evident kinetic energy of large bodies being the degree of organization. If, for example, a vessel of gas is in motion with a certain velocity, we attribute to all the molecules an observable mechanical energy—the common kinetic energy of motion of the centre of gravity. But in addition to this, there is the kinetic energy of the molecules, flying about at random and colliding with the walls, which we term heat. Since disorganization is the essential property of that part of the energy known as heat, one can apply the concepts of the theory of heat only to systems composed of many elementary particles. A single atom cannot be said to have a temperature.

If a quantity of heat represents a definite store of energy, there must be a definite relationship between the calorie and the erg. It has been found experimentally (Joule) that

$$1 \text{ calorie} = 4.184 \times 10^7 \text{ erg}; 1 \text{ kg. cal.} = 4184 \text{ joules.}$$

In what follows we shall assume all energy quantities to be expressed in ergs, so that no conversion factors are required in the formulæ.

We can give another formulation of the First Law—one which is of a negative nature and which is important on account of its immediate intuitive nature and its analogy with the remaining laws. Since the store of energy of an isolated system is constant, *it is impossible to construct a machine which does work without obtaining energy from an external source*. This form of the energy principle is termed the

theorem concerning the impossibility of constructing a *perpetuum mobile* of the first kind.

We must obtain a mathematical expression of the First Law. To this end we introduce the total energy  $U$  as a function of the thermodynamic variables. Evidently this quantity must be a single-valued function of the independent variables, say of  $T$  and  $v$ , for otherwise it would be possible to construct a mechanism which creates energy. For example, if the energy content were greater when 1 litre of air is first warmed from  $0^\circ$  to  $100^\circ$  C. and then compressed to  $10 \text{ cm.}^3$  than when the same amount of air is first compressed to  $10 \text{ cm.}^3$  and then heated to  $100^\circ$  C., it would be possible to gain energy by bringing the substance back to its initial state by the second process, which requires the smaller amount of energy. We must therefore have

$$U = U(T, V) \quad \text{or} \quad U(T, p) \quad \text{or} \quad U(p, V). \quad . \quad . \quad . \quad (1)$$

Hence the differential

$$dU = \left( \frac{\partial U}{\partial T} \right)_v dT + \left( \frac{\partial U}{\partial V} \right)_T dV \quad . \quad . \quad . \quad (2)$$

is the total differential of a function of the thermodynamic variables.

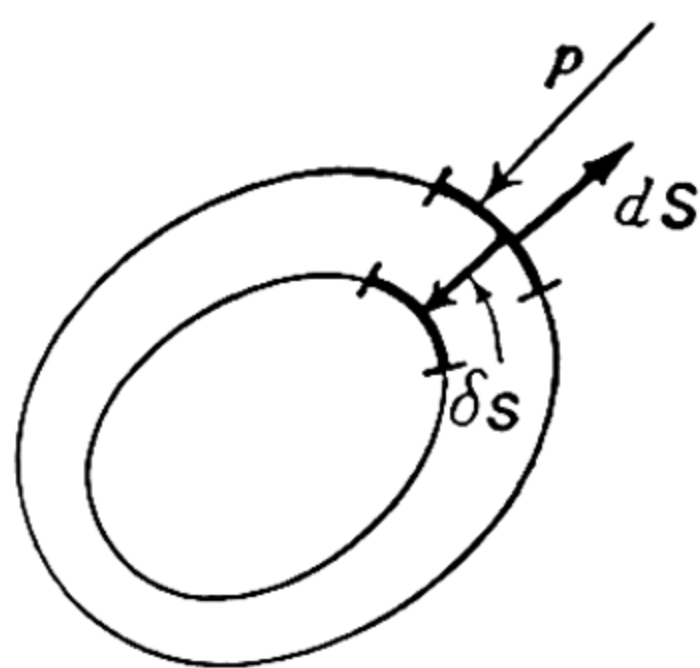


Fig. 1

In contrast with this, the total quantity of heat supplied to a system is not a unique function of the variables, for it is not necessary that the part of the energy of the system which exists in the form of heat be supplied as heat; it may be supplied in part, for example, as mechanical energy in compression. As a result, a small quantity of heat supplied to a system is not the total differential of a function of the thermodynamic variables; it is therefore denoted by  $\delta Q$  instead of by  $dQ$ .

The internal energy of a system may be increased by supplying energy from the outside, part as heat and part by doing mechanical or electrical work. If we take every quantity of energy supplied to be positive, regardless of what form it is in, the First Law is

$$\delta Q + \delta W = dU. \quad . \quad . \quad . \quad . \quad (3)$$

The commonest instance in which work is done is the change of volume accompanying an external pressure  $p$ . If every element  $dS$  of the surface experiences a displacement  $\delta s$ , an amount of work

$$\delta W = - \oint p dS \delta s = -p \oint dS \delta s \quad . \quad . \quad . \quad (4)$$

is done. The negative sign is to be used, since energy is supplied to



the system if  $\delta \mathbf{s}$  forms an obtuse angle with  $d\mathbf{S}$  (cf. fig. 1). Now the integral  $\oint d\mathbf{S} \delta \mathbf{s}$  is merely the change in volume  $dV$  resulting from the displacement. Hence in this special case the First Law may be written

$$\delta Q - p dV = dU, \quad . . . . . (5)$$

or 
$$\delta Q = dU + p dV. \quad . . . . . (5')$$

This last form may also be interpreted thus: the energy supplied to the system in the form of heat goes in part toward increasing the internal energy  $U$  and in part to performing external work.

With regard to the pressure, the following may be added: If the outside pressure  $p$  on the freely movable piston of a gas-filled cylinder is equal to the gas pressure, the piston will remain at rest. If the external pressure is much diminished to a value  $p'$ , a sudden expansion will follow and the gas will acquire energy of flow which, however, is soon converted into heat because of turbulence. In this instance, then, the useful work obtained is smaller than if there had been only an infinitely small difference between inside and outside pressure, and hence an infinitely slow expansion. The latter condition would obviously yield the maximum external work, and we term the process a *reversible* one. Unless otherwise remarked, we shall assume reversibility for all processes discussed.

## 2. Specific Heat at Constant Volume and at Constant Pressure. The Energy Function of a Gas.

According to p. 487, the specific heat per gramme or per mol represents that quantity of heat necessary to impart to one gramme or to one mol a temperature increase of one degree. One must differentiate, however, between the case where heat is supplied at constant volume and that where the pressure is held constant while the heat is supplied. In the first instance no external work is done, while in the second, expansion against the outside pressure does mechanical work. This distinction is important chiefly for gases, on account of their large coefficients of thermal expansion. If we apply the First Law to the energy supplied, the definition of specific heat per mol yields the equation

$$c = \frac{\delta q}{dT} = \frac{du + p dv}{dT}. \quad . . . . . (6)$$

Now the energy  $u$  is a function of the thermodynamic variables, of which we take  $T$  and  $v$  to be the independent ones. We then have for constant volume

$$du = \left( \frac{\partial u}{\partial T} \right)_v dT, \quad . . . . . (7)$$

and the specific heat at constant volume, per mol, is

$$c_v = \left( \frac{\partial u}{\partial T} \right)_v \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot (8)$$

In the second case  $v$  is not constant and

$$\delta q = \left( \frac{\partial u}{\partial T} \right)_v dT + \left[ \left( \frac{\partial u}{\partial v} \right)_T + p \right] dv = c_v dT + \left[ \left( \frac{\partial u}{\partial v} \right)_T + p \right] dv. \quad (9)$$

We now express  $v$  in terms of  $T$  and  $p$  and obtain, for constant pressure,

$$dv = \left( \frac{\partial v}{\partial T} \right)_p dT. \quad \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot (10)$$

Division by  $dT$  yields

$$c_p = c_v + \left[ \left( \frac{\partial u}{\partial v} \right)_T + p \right] \left( \frac{\partial v}{\partial T} \right)_p$$

or

$$c_p - c_v = \left[ \left( \frac{\partial u}{\partial v} \right)_T + p \right] \left( \frac{\partial v}{\partial T} \right)_p \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot (11)$$

This equation takes a particularly simple form for ideal gases. In this case the energy function  $u(T, v)$  may be given explicitly. Since there are no forces acting between the molecules of an ideal gas, the energy must be independent of the distances between molecules, i.e. it must be independent of the volume of the gas. Hence, for an ideal gas

$$\left( \frac{\partial u}{\partial v} \right)_T = 0. \quad \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot (12)$$

This, together with equation (8), gives the energy function of an ideal gas as

$$u = \int_0^T c_v dT + u_0 \cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot (13)$$

Here  $u_0$  is a constant of integration independent of  $T$  and  $v$ —the zero-point energy—which is set equal to zero in the classical thermodynamics. The fact that the internal energy of an ideal gas is independent of the volume is experimentally verified by the expansion experiment of Gay-Lussac. Two large glass bulbs connected by a tube fitted with a valve are immersed in a water bath which serves as a calorimeter. At the start one bulb is evacuated, the other is filled with air. When the valve is opened, the pressures in the bulbs are equalized, but the water bath is found to experience no change in temperature, showing that the internal energy of the gas has not changed after the expansion. It is true that there occurred,

at first, a transformation of heat energy (disorganized energy) into kinetic energy of flow (organized energy), but the latter was converted back into heat when the gas impinged on the walls of the originally empty bulb. In all, the heat energy of the gas remained unaltered.

The equation of state may be expressed as

$$\left(\frac{\partial v}{\partial T}\right)_p = \frac{R}{p}.$$

Using this and (12) in (11) we have

$$c_p - c_v = R \quad . \quad . \quad . \quad . \quad . \quad . \quad (14)$$

for ideal gases. This relationship is known as Mayer's equation.

For a real gas the internal energy is, to a certain extent, dependent upon the volume, for if there are forces of attraction between molecules, work is done against these forces during the expansion. The greater the intermolecular distances, i.e. the greater the volume, the greater is the potential energy corresponding to these forces. Using van der Waals' equation, it is easy to obtain an expression for the dependence of internal energy upon volume. If there is a small positive change of volume  $dv$ , an amount of work

$$dW = + \frac{a}{v^2} dv \quad . \quad . \quad . \quad . \quad . \quad . \quad (15)$$

is done against the cohesive forces. Since the added term for real gases is always small, we set the internal energy equal to the sum of a term depending only on the temperature and one depending only on the volume:

$$u = u_1(T) + u_2(v), \quad \text{i.e.} \quad \left(\frac{\partial u}{\partial v}\right)_T = \frac{du_2}{dv}. \quad . \quad . \quad (16)$$

If no energy is supplied from without, the change of the second term is equal to the work done on the system:

$$du_2 = \frac{du_2}{dv} dv = +dW = + \frac{a}{v^2} dv;$$

hence

$$u_2 = - \frac{a}{v}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (17)$$

and since

$$\left(\frac{\partial u}{\partial T}\right)_v = c_v,$$

$$u = \int_0^T c_v dT - \frac{a}{v} + u_0. \quad . \quad . \quad . \quad . \quad (18)$$



If the specific heat is independent of the temperature (this is true about as far as  $1000^{\circ}$  abs. for all gases except hydrogen), then

$$u = c_v T - \frac{a}{v} + u_0. \quad . \quad . \quad . \quad . \quad (18')$$

The constant  $u_0$  is subject to the same remarks as in the case of an ideal gas (p. 483).

For a van der Waals gas we then have

$$c_p - c_v = \left( p + \frac{a}{v^2} \right) \left( \frac{\partial v}{\partial T} \right)_p. \quad . \quad . \quad . \quad . \quad (19)$$

Differentiation of the equation of the state gives

$$\left( \frac{\partial v}{\partial T} \right)_p = \frac{R'}{p + \frac{a}{v^2} - \frac{2a}{v^3} (v - b)}.$$

Putting this value in equation (19) and using the equation of state, we obtain

$$c_p - c_v = \frac{R'}{1 - \frac{2a(v-b)^2}{v^3 R' T}},$$

or, replacing  $R'$  by  $R$ , we have approximately, since  $a$  and  $b$  are small,

$$c_p - c_v \approx R \left( 1 + \frac{2a}{vRT} \right) \quad . \quad . \quad . \quad . \quad (20)$$

for a van der Waals gas.

### 3. Adiabatic Change.

So far we have considered processes in which one of the primary variables  $T$ ,  $p$  or  $v$  was held constant. Instead, a process may be characterized by holding constant *any function* of the thermodynamic variables. Adiabatic processes are of this kind. The function which is held fixed in this process is the entropy; this function will be discussed in the next chapter. Adiabatic processes are, as the name indicates, those in which no heat is communicated to or abstracted from the system. This condition may be realized in practice in two ways: either the entire system is enclosed in a heat insulating envelope or else the processes take place so rapidly that there can be no appreciable heat transferred by conduction even when the heat insulation is imperfect.

The First Law furnishes at once a differential equation for the adiabatic process:

$$du + p dv = 0. \quad . \quad . \quad . \quad . \quad . \quad (21)$$

If  $u$  is expressed in terms of  $p$  and  $v$ , this equation represents a family of curves known as the adiabatic lines or simply *adiabatics*. The equation of the adiabatics is especially simple for ideal gases: since we have

$$du = c_v dT \quad \text{and} \quad dT = \frac{p dv + v dp}{R},$$

and since  $R = c_p - c_v$ , (21) becomes

$$\frac{c_p}{c_v} \frac{dv}{v} + \frac{dp}{p} = 0. \quad . \quad . \quad . \quad . \quad . \quad (22)$$

Using the symbol  $\gamma$  for the ratio  $c_p/c_v$  and integrating, there results

$$pv^\gamma = p_0 v_0^\gamma = \text{const.} \quad . \quad . \quad . \quad . \quad . \quad (23)$$

Since  $\gamma > 1$ , the adiabatics are always steeper than the isotherms ( $pv = \text{const.}$ ) in the  $pv$ -diagram. If we represent the relationship between  $T$ ,  $p$  and  $v$  in a rectangular system of co-ordinates, equation (23) represents the projection on the  $pv$ -plane of a space curve lying on the surface  $pv = RT$ . Making use of the equation of state, we find that the other projections are given by

$$Tv^{\gamma-1} = T_0 v_0^{\gamma-1} = \text{const.} \quad . \quad . \quad . \quad . \quad (24)$$

and 
$$Tp^{(1-\gamma)/\gamma} = T_0 p_0^{(1-\gamma)/\gamma} = \text{const.} \quad . \quad . \quad . \quad (25)$$

Thus  $\gamma$  may be determined for an ideal gas from the adiabatic change in volume. A method which is more general, since it is not restricted to ideal gases, is that which employs the velocity of sound in the gas. The relationship

$$v = \sqrt{\frac{\text{modulus of elasticity}}{\text{density}}} \quad . \quad . \quad . \quad . \quad (26)$$

given on p. 183 for longitudinal waves in solid bars is valid also for fluids if we use the volume modulus of elasticity (see p. 213). If we insert the numerical values of density and modulus of elasticity for air, which (*Ex.* 119, p. 500) is equal to the pressure, for an ideal gas, we obtain a value of  $v$  which does not agree with observation. Newton was the first to compute the velocity of sound in this way. The source of the error in his result was pointed out by Laplace. It is not that air was incorrectly assumed to be an ideal gas; the discrepancy is of a much greater order of magnitude. The true reason is that the compression take place so rapidly in a sound wave that the resulting heat does not have time to dissipate itself; hence the process is adiabatic, while the modulus of elasticity in the example, p. 500, assumes that the

temperature is constant. Now it is readily shown that the adiabatic modulus of elasticity is equal to  $\gamma$  times the isothermal modulus. To prove this, take  $p$  and  $v$  as independent variables. The differential equation of the adiabatics is then, by (5) (p. 503),

$$\left(\frac{\partial u}{\partial p}\right)_v dp + \left[\left(\frac{\partial u}{\partial v}\right)_p + p\right] dv = 0. \quad \dots \quad (27)$$

On account of equation (8) (p. 504), this gives

$$\begin{aligned} \left(\frac{\partial p}{\partial v}\right)_{\text{ad}} &= - \frac{p + \left(\frac{\partial u}{\partial v}\right)_p}{\left(\frac{\partial u}{\partial p}\right)_v} = - \frac{p + \left(\frac{\partial u}{\partial T}\right)_p \left(\frac{\partial T}{\partial v}\right)_p}{\left(\frac{\partial u}{\partial T}\right)_v \left(\frac{\partial T}{\partial p}\right)_v} \\ &= - \frac{\left(\frac{\partial T}{\partial v}\right)_p \left[ p \left(\frac{\partial v}{\partial T}\right)_p + \left(\frac{\partial u}{\partial T}\right)_p \right]}{c_v \left(\frac{\partial T}{\partial p}\right)_v}. \quad \dots \quad (28) \end{aligned}$$

The existence of an equation of state connecting the variables, say  $f(p, v, T) = 0$ , implies

$$\frac{\partial f}{\partial p} dp + \frac{\partial f}{\partial v} dv + \frac{\partial f}{\partial T} dT = 0,$$

so that, as at (5) (p. 494)

$$\frac{\left(\frac{\partial T}{\partial v}\right)_p}{\left(\frac{\partial T}{\partial p}\right)_v} = - \left(\frac{\partial p}{\partial v}\right)_T. \quad \dots \quad (29)$$

Thus we need only show that the bracketed expression in (28) is equal to  $c_p$  in order to prove our statement. If we now take  $T$  and  $p$  as independent variables, we have, by the First Law, at constant pressure,

$$\delta q = \left(\frac{\partial u}{\partial T}\right)_p dT + p \left(\frac{\partial v}{\partial T}\right)_p dT,$$

whence

$$c_p = \frac{\delta q}{dT} = p \left(\frac{\partial v}{\partial T}\right)_p + \left(\frac{\partial u}{\partial T}\right)_p. \quad \dots \quad (30)$$

Thus (28) becomes

$$\left(\frac{\partial p}{\partial v}\right)_{\text{ad}} = \gamma \left(\frac{\partial p}{\partial v}\right)_T, \quad \dots \quad (31)$$



and so, from the definition of  $E$ ,

$$E_{ad} = \gamma E_{is}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (31')$$

The value of the adiabatic modulus is obtained from the speed of sound, while static experiments give the isothermal modulus. Hence  $\gamma$  may be computed.

An adiabatic process which has found an important technical application in the air liquefying machine (Linde) is the so-called throttled expansion of a real gas. In the Gay-Lussac expansion experiment (p. 504) the gas in the first bulb attained kinetic energy of flow and was at the same time cooled; the energy was again converted into heat, however, when the motion was arrested in the second vessel, so that within the somewhat large limits of error no net change in temperature was observed. The accuracy of observation is increased very much if the expansion is regulated so that no energy is converted into energy of flow. This is the so-called "porous plug" experiment of Joule and Thomson (Kelvin). With this arrangement it is not necessary to wait for the system to come to equilibrium; rather, the temperature of the expanded gas itself may be measured. Since this gas has received no appreciable kinetic energy, its temperature is a measure of the change of internal energy accompanying the increase in volume. The apparatus is represented diagrammatically in fig. 2.

By continuous motion inward of the piston 1 and continuous withdrawal of piston 2 the pressure  $p_1$  is maintained on the left, the pressure  $p_2$  on the right. In

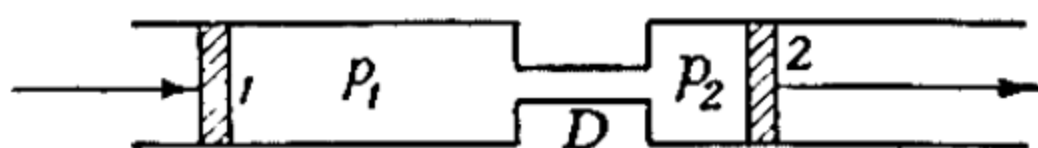


Fig. 2

the original experiments the throttle  $D$  was a plug of cotton or silk. If one mol of gas is forced through in this manner, the amount of work done on the left is  $p_1 v_1$ , while the gas has done an amount of external work  $p_2 v_2$  on the other side. The increase in the internal energy of this mol of gas is then the excess of the applied work over the work obtained from the system:

$$u_2 - u_1 = p_1 v_1 - p_2 v_2$$

or

$$u_2 + p_2 v_2 = u_1 + p_1 v_1. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (32)$$

In this process, then, the "heat function" or *enthalpy*, or *total heat*  $h = u + pv$  remains constant. Since each term is a function of the thermodynamic variables, the enthalpy is also a function of these variables. The change in temperature  $\Delta T$  accompanying a change in volume  $\Delta v$  is readily given for a van der Waals gas:

$$h = u_0 + c_v T - \frac{a}{v} + \left( \frac{R'T}{v-b} - \frac{a}{v^2} \right) v, \quad . \quad . \quad (33)$$

whence it follows that

$$\begin{aligned}\Delta h = 0 &= c_v \Delta T + \frac{2a}{v^2} \Delta v + \frac{R'v}{v-b} \Delta T - \frac{R'Tv}{(v-b)^2} \Delta v + \frac{R'T}{v-b} \Delta v \\ &= \left( c_v + \frac{R'v}{v-b} \right) \Delta T + \left( \frac{2a}{v^2} - \frac{R'Tb}{(v-b)^2} \right) \Delta v. \quad \cdot \cdot \cdot (34)\end{aligned}$$

Neglecting  $b$  as compared with  $v$  and making the further approximation of replacing  $R'$  by  $R$  and  $c_v + R$  by  $c_p$ , we obtain, after solving for  $\Delta T$ :

$$\Delta T = \frac{\Delta v}{v^2 c_p} (RTb - 2a). \quad \cdot \cdot \cdot \cdot \cdot (35)$$

In this experiment the temperature difference that arises represents the difference between two opposing processes. Expansion makes the attraction between molecules produce cooling, since work must be done against these forces. On the other hand, the space occupied by the molecules, which is characterized by  $b$ , is responsible for a warming effect when the gas expands.

Of greater importance is a knowledge of the change in temperature as a function of the pressure change  $\Delta p$ . To obtain an expression for  $v$  in terms of  $p$  which is accurate to the same order as (35), we may replace  $v$  by the equivalent expression  $RT/p$  from the ideal gas equation and obtain

$$\Delta T \approx \frac{\Delta p}{c_p} \left( \frac{2a}{RT} - b \right). \quad \cdot \cdot \cdot \cdot \cdot (36)$$

Since  $\Delta p < 0$ , there will be a decrease in temperature as long as

$$T < \frac{2a}{Rb}.$$

The numerical values for air are such that a considerable cooling effect is obtained even at  $0^\circ \text{C.}$ , and this effect is employed in the Linde machine to cool the following masses of air. For hydrogen, however, there is a temperature rise when the exit temperature is  $0^\circ \text{C.}$ , hence this gas must be pre-cooled with liquid air.

In reality, the approximations made above are quite crude in the neighbourhood of the liquefaction point. If (34) is solved for  $\Delta T$  without neglecting any terms, the factor of  $\Delta v$  when set equal to zero no longer yields a fixed transition temperature but an equation connecting  $T$  and  $v$ —a transition curve. This shows that even at lower temperatures there are pressures at which the sign of the effect reverses (see following exercise).

*Ex. 120.* Find the transition curve in terms of reduced variables.

#### 4. Application of the First Law to Thermochemistry.

In all chemical reactions we observe a heat exchange, i.e. the reaction is accompanied either by the evolution of heat—a so-called

exothermal reaction—or heat is absorbed in the process—an endothermal reaction. Most chemical changes are of the former type. If the reaction takes place in a calorimeter, the heat of reaction may be measured. Contrary to our former practice of counting heat supplied to a system as positive, the heat of reaction in thermochemistry is taken to be positive for exothermal reactions. This will be expressed by representing the reaction heat by  $Q^- = -Q$ ;  $Q^-$  is measured in thermal units (calories), and is read “ $Q$  minus”.

If the reaction takes place without doing external work (i.e. for gases, at constant volume), the heat of reaction gives the change of energy of the system resulting from the reaction. On the other hand, if the process takes place at constant pressure, then—practically only in the case of gases—the external work must also be taken into account. If the volume of the system is  $V_1$  before the occurrence of the reaction and  $V_2$  after, this work amounts to  $p(V_2 - V_1)$ . Designating the two heats of reaction by the subscripts  $v$  and  $p$ , as done above for the specific heats, we have, at constant volume,

$$Q^-_v = U_1 - U_2. \quad . \quad . \quad . \quad . \quad . \quad (37)$$

At constant pressure,

$$\text{or} \quad \left. \begin{aligned} Q^-_p &= U_1 - U_2 - p(V_2 - V_1) \\ Q^-_p &= U_1 + pV_1 - (U_2 + pV_2) = W_1 - W_2 \end{aligned} \right\}. \quad (38)$$

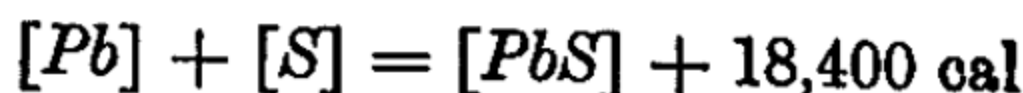
Thus the heat exchange at constant pressure gives the change in enthalpy  $H$  before and after the reaction. Since it is sufficiently accurate to consider the gases ideal in thermochemistry, we have

$$pV_1 = n_1RT$$

$$\text{and so} \quad Q^-_p = Q^-_v + (n_1 - n_2)RT = Q^-_v + \Delta nRT, \quad . \quad (39)$$

where  $\Delta n$  is the difference in the number of mols of gaseous components of the system before and after the reaction.

For reactions at constant pressure, the enthalpy is the same as the energy content; moreover, in no case do these two quantities differ much. We adopt the following notation for thermochemical equations: The enthalpy  $H$  of one mol of a substance in the liquid state is denoted by the chemical symbol written without brackets, for the solid state by the symbol in square brackets, and for the gaseous state in parentheses. Thus the thermochemical equation



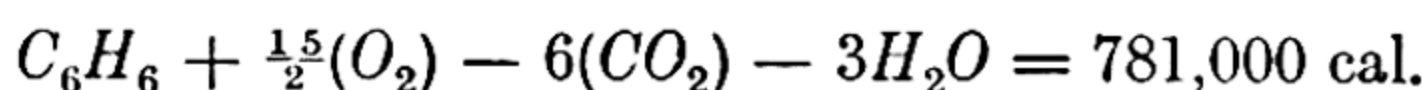
$$\text{or} \quad [Pb] + [S] - [PbS] = 18,000 \text{ cal}$$

means that one mol of solid lead and one mol of solid sulphur together have an energy content 18,400 cal larger than one mol of lead sulphide.

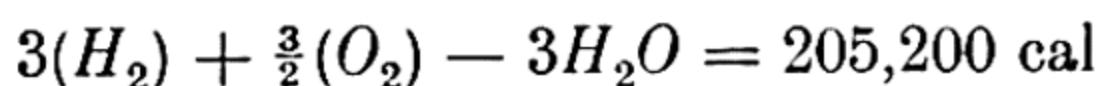
The main significance of the First Law is that energy content



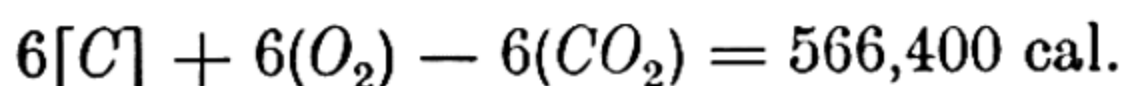
and enthalpy are functions only of the thermodynamic variables; hence the energy differences must always be the same, whether the reaction consists of a single process or involves intermediate steps. This fact was expressed by Hess before the First Law was known. Hess's Law enables us to combine thermochemical equations; this often makes possible the calculation of the heat of reaction in cases where it would be extremely difficult to measure, or for reactions which are purely hypothetical. In organic chemistry, for example, the heat of formation of a complicated compound is determined by combustion, first of the compound and then of the constituents. The combustion usually takes place in a bomb calorimeter at constant volume. In order to show that this method yields a sufficient number of equations for calculating the heat of formation we consider the formation of liquid benzene from solid carbon and gaseous hydrogen, the process occurring at constant pressure. From the reaction in a bomb calorimeter (constant volume) we find



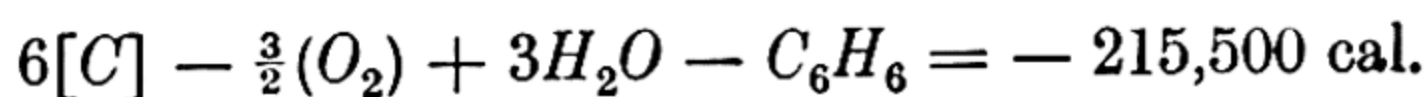
Since the data on the remaining reactions refer to constant pressure conditions, we convert the above reaction to constant pressure also. In the process under consideration,  $3/2$  mol of gas vanishes, and since  $R$  is approximately equal to 2 cal, we must add 900 cal to the right member of the preceding equation, assuming room temperature to be about  $300^\circ$  abs. This quantity of heat is relatively small compared with the total amount, showing—as indicated above—that no great error is involved in replacing  $Q_p^-$  by  $Q_v^-$ , especially since the heat of reaction cannot be measured with any great accuracy. In addition, we have the following data for the combustions at constant pressure:



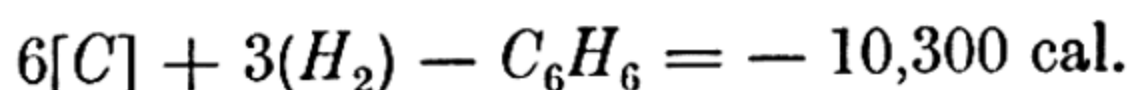
and



Increasing the right member of the first of these three equations by the required 900 cal and subtracting it from the third we obtain



Adding the second equation to this,



Thus the heat of formation of benzene from carbon and hydrogen is negative—energy must be supplied to make the reaction occur. It must be mentioned, however, that the heat of formation is not the correct criterion for the occurrence of the process, although in general—as in the present instance—exothermal reactions are easier to pro-

duce than are endothermal. This is known as Berthelot's Principle.

The First Law of Thermodynamics enables us to find the law of dependence of heat of reaction on temperature. For a reaction taking place at constant volume,

$$\nu_a a + \nu_b b + \dots = \nu_c c + \nu_d d + \dots, \quad (40)$$

where  $\nu_a, \nu_b, \dots$  are the numbers of mols, and  $a, b, \dots$  are the formulæ of the substances entering into the reaction. At a temperature  $T$ ,

$$(\nu_a a)_T + (\nu_b b)_T + \dots - (\nu_c c)_T - (\nu_d d)_T - \dots = q_v^-(T), \quad (41)$$

and at a temperature  $T + dT$ ,

$$\begin{aligned} &(\nu_a a)_{T+dT} + (\nu_b b)_{T+dT} + \dots - (\nu_c c)_{T+dT} - (\nu_d d)_{T+dT} - \dots \\ &= q_v^-(T) + \frac{dq_v^-}{dT} dT. \end{aligned} \quad (42)$$

Now at constant volume

$$du = c_v dT, \quad (43)$$

so that

$$\begin{aligned} &(\nu_a a)_T + \nu_a c_{va} dT + (\nu_b b)_T + \nu_b c_{vb} dT + \dots \\ &- (\nu_c c)_T - \nu_c c_{vc} dT - (\nu_d d)_T - (\nu_d c_{vd}) dT - \dots = q_v^-(T) + \frac{dq_v^-}{dT} dT, \end{aligned} \quad (44)$$

or 
$$\frac{dq_v^-}{dT} = \sum \nu_i c_{vi}. \quad (44')$$

In this summation, the number of mols is to be counted positive before the reaction and negative afterward. Exactly the same equations result for reactions at constant pressure, only it is necessary to remember that the heat is supplied at constant pressure, so that  $c_v$  is to be replaced by  $c_p$ :

$$\frac{dq_p^-}{dT} = \sum \nu_i c_{pi}. \quad (45)$$

Over a small temperature interval,  $c_v$  or  $c_p$  may be considered constant; this is no longer allowable when large intervals are considered. For example, if we wish to introduce the heats of reaction at absolute zero instead of those observed at room temperature, it is necessary to take

$$q_v^-(T) = q_{v0}^- + \sum \nu_i \int_0^T c_{vi} dT \quad \text{and} \quad q_p^-(T) = q_{p0}^- + \sum \nu_i \int_0^T c_{pi} dT. \quad (46)$$

These integrals may be evaluated if the variation of the specific heats is known.

## CHAPTER XXXI

### THE SECOND LAW OF THERMODYNAMICS: THE LAW OF ENTROPY

#### 1. The Carnot Cycle and the Ideal Heat Engine.

The First Law of Thermodynamics makes a statement concerning the energy balance which must be maintained in any transformation between mechanical or electromagnetic energy and heat. This law, however, does not tell whether or not this transformation will take place under given conditions, nor does it say what portion of the energy will be transformed. Even before the First Law was discovered it was known that any amount of mechanical energy could be completely converted into heat at will, e.g. by friction, but only a partial conversion of heat into mechanical energy could be realized with the aid of heat engines. By studying these relationships, Carnot\* was able to represent the processes taking place in a heat engine by means of a cycle which, as we shall find, represents the best possible heat engine. This ideal machine makes use of two large reservoirs of heat which are at temperatures  $T_1$  and  $T_3$  respectively, with  $T_1 > T_3$ .† The working substance may be taken to be one mol of an ideal gas enclosed in a cylinder equipped with a frictionless piston. By means of a frictionless guiding mechanism of some kind, the cylinder may be placed in thermal contact with each of the reservoirs in turn, or may be completely insulated from them. We provide, further, that the processes which the working substance undergoes be reversible. By this we mean the following. If the gas is allowed to expand by decreasing the external pressure, this must not be done by a sudden, finite change of pressure, but in such a way that the external pressure differs only infinitesimally from the internal; thus the piston receives no appreciable kinetic energy which would require to be removed by braking. Under such conditions the process may be carried out equally well in the reverse order, provided that the external pressure is a little greater than the internal. In this case the work done differs only in algebraic sign from that done in the expansion process. It is evident

\* S. Carnot, *Réflexions sur la puissance motrice du feu, et sur les machines propres à développer cette puissance*, Paris (1824); Eng. trans. by R. H. Thurston (1890).

† These heat containers are sometimes referred to as "source" and "sink" or as "boiler" and "condenser" respectively.



that the rapidly running engines used in practice are very different from the ideal Carnot mechanism.

A *cyclic process* or *cycle* is a process in which the machine is again brought back to its initial state; this is a necessary property of a machine that is to run arbitrarily long. Let us cause the working substance to pass through the following cycle (fig. 1):

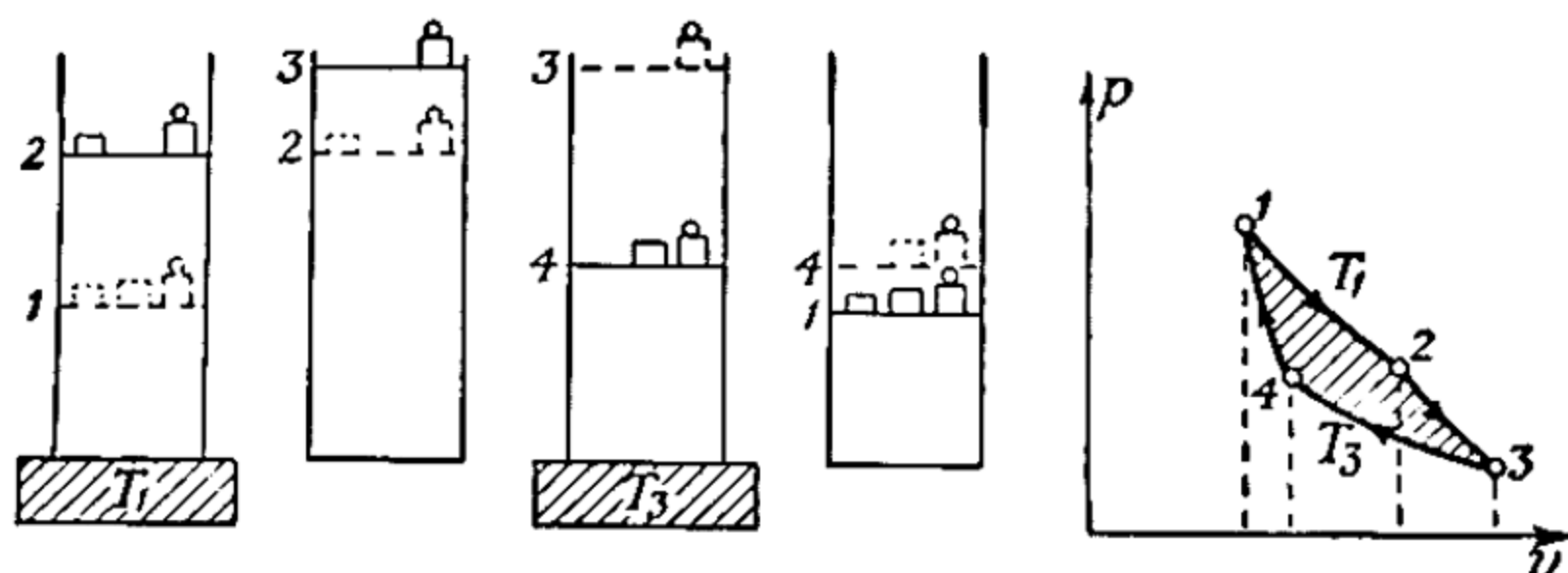


Fig. 1

1. Connect the cylinder with the source of heat at temperature  $T_1$  (the boiler) and reduce the external pressure (e.g. by removing weights from the piston), thus allowing the gas to expand. In doing so, the gas does work against the external pressure and the equivalent amount of energy is taken from the boiler. Since we assumed the boiler to be very large, we need not consider its temperature drop during the process and so may take the change in volume of the working substance to be isothermal. If the gas expands from the volume  $v_1$  to  $v_2$  the external work done is \*

$$W_{12}^- = \int_{v_1}^{v_2} p dv = RT_1 \int_{v_1}^{v_2} \frac{dv}{v} = RT_1 \log \frac{v_2}{v_1}. \quad \dots (1)$$

2 The cylinder is now removed from the boiler and insulated. The gas is then allowed to expand adiabatically until its temperature has dropped to  $T_3$ , that of the condenser. By equation (24) (p. 507) the volume  $v_3$  at that temperature is given by

$$v_3 = v_2 \left( \frac{T_3}{T_2} \right)^{1/(1-\gamma)} = v_2 \left( \frac{T_3}{T_1} \right)^{1/(1-\gamma)} \quad \dots (2)$$

since  $T_2 = T_1$ . Thus the work done in this adiabatic expansion is

$$\begin{aligned} W_{23}^- &= \int_{v_2}^{v_3} p dv = p_2 v_2^\gamma \int_{v_2}^{v_3} \frac{dv}{v^\gamma} \\ &= \frac{p_2 v_2^\gamma}{1-\gamma} (v_3^{1-\gamma} - v_2^{1-\gamma}) = \frac{RT_2}{1-\gamma} \left\{ \left( \frac{v_3}{v_2} \right)^{1-\gamma} - 1 \right\}, \quad \dots (3) \end{aligned}$$

\* As in the case of quantity of heat, the work done on a system is reckoned positive. When, infrequently, we count the work done by a system as positive we denote it by  $W^-$  ("W minus"), i.e.  $W^- = -W$ .

or, by (2), this is equal to

$$\frac{RT_2}{1-\gamma} \left( \frac{T_3}{T_2} - 1 \right) = \frac{R}{\gamma-1} (T_2 - T_3) = \frac{R}{\gamma-1} (T_1 - T_3).$$

3. The cylinder is now placed in thermal contact with the condenser (at temperature  $T_3$ ) and the gas is isothermally compressed to a volume  $v_4$  such that if it be adiabatically compressed from this point to the initial volume  $v_1$  it attains the initial temperature  $T_1$ . This volume is given by

$$v_4 = v_1 \left( \frac{T_4}{T_1} \right)^{1/(1-\gamma)} = v_1 \left( \frac{T_3}{T_1} \right)^{1/(1-\gamma)} \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad (4)$$

since  $T_4 = T_3$ . The work done in this compression is negative, since it is really supplied to the system. Its magnitude is

$$W_{34}^- = \int_{v_4}^{v_3} p dv = RT_3 \log \frac{v_4}{v_3} \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad (5)$$

The heat generated is delivered to the condenser.

4. The cylinder is removed from the condenser, is thermally insulated and compressed adiabatically to  $v_1$ . An amount of work

$$W_{41}^- = \int_{v_4}^{v_1} p dv = \frac{R}{\gamma-1} (T_4 - T_1) = \frac{R}{\gamma-1} (T_3 - T_1) \quad \cdot \quad (6)$$

is thus done. The same remarks apply to the algebraic sign as in the third phase of the process.

The machine itself is now in its initial configuration. A certain quantity of thermal energy has been abstracted from the boiler (temperature  $T_1$ ) and delivered to the condenser (temperature  $T_3$ ) and, in addition, a certain amount of mechanical work has been done. This is given by the sum

$$W_{12}^- + W_{23}^- + W_{34}^- + W_{41}^-,$$

and is represented in the so-called *indicator diagram* (fig. 1) by the area enclosed by the path covered in the process. Hence external work is done when the path is traversed in such way that the enclosed area is always on the right, as is evident at once by considering the expansion from  $v_1$  to  $v_2$ , in which the work done is represented by the area between the curve  $p = f(v)$ , the axis of abscissæ, and the lines through points 1 and 2 drawn parallel to the  $p$ -axis. Adding the amounts of work done in the individual steps, one sees at once that the amounts along the adiabatics cancel out and there remains

$$W^- = RT_1 \log \frac{v_2}{v_1} + RT_3 \log \frac{v_4}{v_3} \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad (7)$$

But it follows from equations (2) and (4) that

$$\frac{v_4}{v_3} = \frac{v_1}{v_2} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

Hence

$$W^- = R (T_1 - T_3) \log \frac{v_2}{v_1} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (9)$$

A quantity of heat

$$Q_1 = RT_1 \log \frac{v_2}{v_1}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (10)$$

equivalent to the work done in the first step of the process, was removed from the boiler and delivered to the engine. Of this the fraction

$$\eta = \frac{W^-}{Q_1} = \frac{T_1 - T_3}{T_1} = 1 - \frac{T_3}{T_1} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (11)$$

was converted into useful mechanical work, while the quantity of heat

$$Q_3 = RT_3 \log \frac{v_2}{v_1} = Q_1 \frac{T_3}{T_1} = Q_1 (1 - \eta) \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (12)$$

was returned to the condenser. The fraction  $\eta$  is called the *efficiency* of the heat engine. Since all the steps of the process are reversible, the engine may be used also to extract a quantity of heat  $Q_3$  from the condenser, delivering it to the boiler. This requires the expenditure of an amount of work

$$W = \eta Q_3 = \frac{\eta Q_3}{1 - \eta} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (13)$$

The efficiency of the Carnot engine is the highest attainable by any means. To show this, assume that there exists a "super-Carnot engine" with still higher efficiency  $\bar{\eta}$ . By coupling this machine to a Carnot engine we would have a heat engine which—while it would not be a perpetual motion machine, since it does not violate the First Law—would nevertheless be equivalent to one from the practical standpoint. Such a mechanism is called a *perpetuum mobile of the second kind*. Let the super-Carnot engine take a quantity of heat  $Q_\infty$  from the boiler, converting  $\bar{\eta}Q_\infty$  of this into useful work and returning  $(1 - \bar{\eta})Q_\infty$  to the condenser. This quantity of heat, however, may be recovered immediately by means of a Carnot engine; by equation (13), this requires an amount of work

$$W = \frac{1 - \bar{\eta}}{1 - \eta} \eta Q_\infty$$



In all, the combination has done an amount of work

$$W^-_x = \bar{\eta}Q_x - \frac{1 - \bar{\eta}}{1 - \eta} \eta Q_x = \frac{(\bar{\eta} - \eta)Q_x}{1 - \eta}, \quad \cdot \cdot \quad (14)$$

which is positive. The thermal equivalent of this work has been abstracted from the boiler without the delivery of any of it to the condenser, which therefore is not required at all. This means that the mechanism would do mechanical work and would cool the boiler without causing any alteration whatsoever in the condition of any other body. Since enormous sources of heat are available (e.g. in the oceans), a machine of this kind would be fully as important as a perpetual motion machine. Experience shows, however, that it is impossible to construct a machine of this kind, and this is precisely what the Second Law of Thermodynamics says.

Before we formulate this law, it must be mentioned that a satisfactory definition of the temperature scale is given by equation (11) (p. 517). Thus, in order to compare the temperatures of two sources of heat, we use them as boiler and condenser of a Carnot engine and determine its efficiency. The condenser is at the absolute zero of temperature if the efficiency is unity. This defines the *thermodynamic* or *Kelvin temperature scale*, which is identical with that based on the equation of state of an ideal gas, as is evident from the derivation. By means of suitable thermodynamic reasoning and measurements, any thermometer may be reduced to the thermodynamic scale. However, the details cannot be given here. At not too low temperatures the readings of a helium thermometer agree with the thermodynamic scale.

*Ex. 121.* Is the heat developed in electric resistors an economical way of heating a building?

## 2. Formulation and Interpretation of the Second Law.

In conformity with the result of the previous section we give the Second Law the formulation due to Thomson and Planck:

*It is impossible to construct a continuously operating machine which does mechanical work and which cools a source of heat without producing any other effects.*

This is a fact of experience, just as is the First Law, and can be confirmed or disproved only by experiment. We shall find that the consequences of the Second Law are exceedingly numerous and that the experimental confirmation of these consequences is at the same time a verification of the validity of the Second Law, which, moreover, has a logical interpretation from the atomistic point of view.

The first consequence is that certain energy transformations involving heat are irreversible. A case in point is the conversion of work into heat by means of friction. If, for example, a vessel containing

a rotating liquid is left to itself the liquid will soon come to rest, since its kinetic energy is soon converted into heat; this causes a slight temperature rise which may be measured. It would be an event absolutely unheard of, however, if a liquid at rest were to start rotating, its temperature dropping at the same time, although this would not be a violation of the First Law. If the process opposite to the generation of heat by friction existed, it would have exactly the property of a *perpetuum mobile* of the second kind. If, then, we assume the Second Law to be valid, such a process does not exist.

Another irreversible process is the isothermal expansion of an ideal gas without the performance of external work. This is the process we met in the Gay-Lussac experiment (p. 504). If it were possible to compress a gas isothermally without using external work, we could construct a *perpetuum mobile* of the second kind by allowing a gas to expand isothermally and do external work, the required energy being taken from a reservoir in the form of heat; then, using the previously mentioned process, the gas could be compressed again to its original volume without requiring any work.

Expansion without external work is represented also by the diffusion of two gases. If we remove a partition separating two gases in a vessel, each gas ultimately fills the entire vessel, since the entire space is available to the molecules of either gas. According to the Second Law, this process cannot be reversed without doing work. In this process and in the one discussed in the preceding paragraph heat enters only indirectly, since no changes in temperature occur. Nevertheless, the phenomenon of diffusion throws light on the nature of the Second Law. We may represent the circumstances by a model as follows. Place a large number of white balls, representing the molecules of the first gas, in the bottom of an urn. Above these place a layer of coloured balls, representing the molecules of the second gas. If the urn is now shaken, the two kinds of spheres will mix. It is improbable to the highest degree—and the improbability increases with the number of balls used—that the shaking will again produce a configuration in which all the coloured balls are again on top. To be sure, there is no physical law forbidding such a separation, just as there is no real reason why a number of pieces of type thrown at random from a box should not arrange themselves so as to spell out a given sentence, but the Theory of Probability shows that such a “self-arrangement” is *practically* out of the question. The remarkable one-sidedness of the energy changes mentioned above is to be interpreted atomistically by the fact that the change takes place from a state of order to one of disorder, but not the other way.

We now endeavour to give the Second Law a mathematical formulation from which further consequences may be deduced mathematically. The First Law characterizes the energy  $U$  which depends only



upon the thermodynamic variables. Is it also possible to find a characteristic function for the Second Law which has this property? The Carnot cycle gives us a suggestion. After completion of the cycle, i.e. when the system is again in its initial configuration, we have by equations (10) and (12) (p. 517),

$$\frac{Q_1}{T_1} = \frac{Q_3}{T_3} \text{ or } \frac{Q_1}{T_1} + \frac{Q_3}{T_3} = 0. \quad \cdot \cdot \cdot (15)$$

We wish to generalize this relationship so that it applies to any cyclic process. Imagine such a process carried out with an arbitrary system. The cycle is represented by a closed curve in the  $pV$ -diagram; the direction in which this curve is to be described is that corresponding

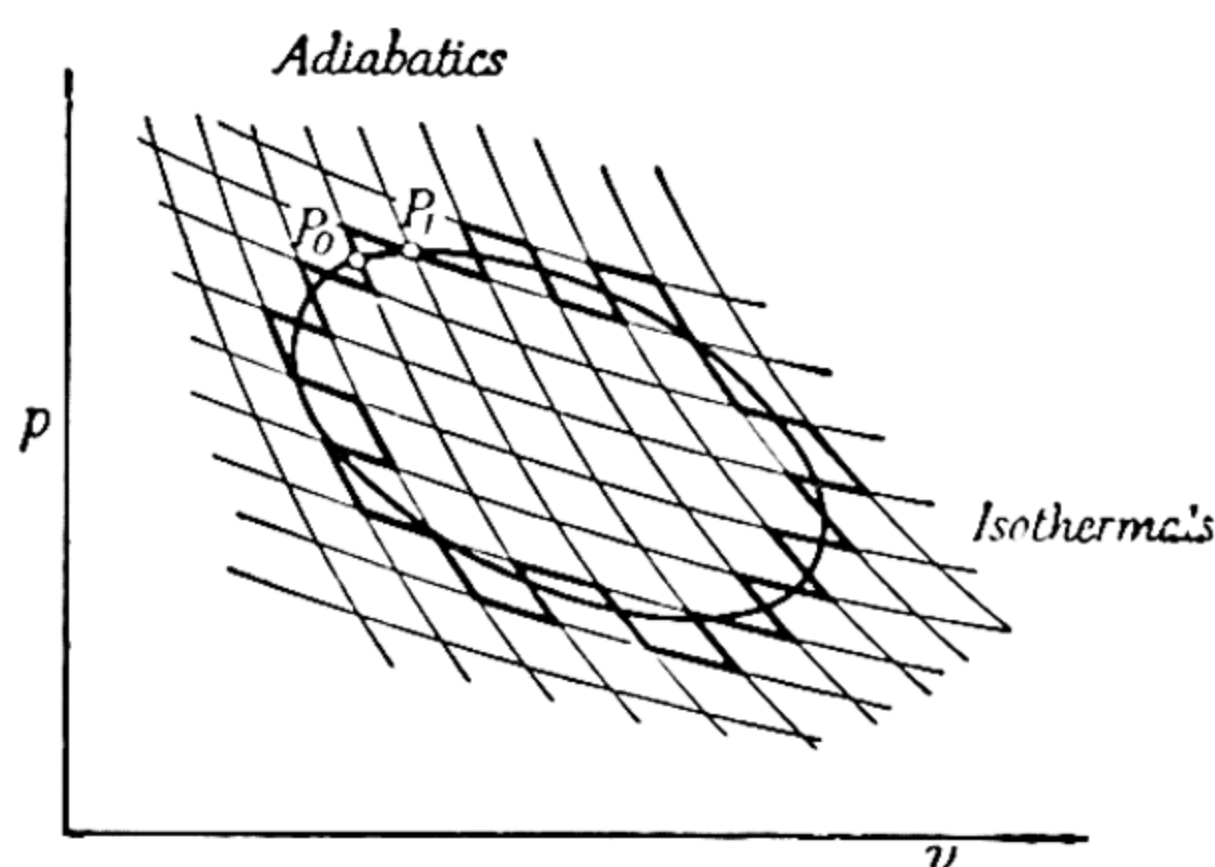


Fig. 2

to external work done by the system. Imagine, now, the  $pV$ -plane to be covered by a fine network of adiabatics and isothermals (fig. 2). If the network is taken arbitrarily fine, the work done may be represented by the area enclosed by the zigzag line instead of that bounded by the curve itself. The former path is traced by moving from a point  $P_0$  of the curve to a neighbouring point  $P_1$  along an adiabat as far as the intersection with the isothermal through  $P_1$ , and then along this isothermal to  $P_1$ , instead of moving directly from one point to the other along the original curve.

We specify, in addition, that all changes are to be reversible in the sense used in connexion with the Carnot cycle. Further, assume that the heat supplied or obtained along the isothermals is not exchanged with the boiler or condenser directly, but that the work which can be got from it is really obtained with the help of a Carnot engine operating between a body at temperature  $T_0$  and the system itself. The Carnot engine delivers to the boiler the heat generated along the isothermals, external work being used up in the process. Since only a small segment of each isothermal is traversed, the individual quantities of heat supplied to the system are small and so are



denoted by  $\Delta Q_\sigma$ . The quantities of heat which the Carnot engine abstracts from the boiler at temperature  $T_0$  are denoted by  $\Delta Q_0^-$ . Part of this heat is converted into mechanical work, while  $\Delta Q_\sigma$  is delivered to the system in question. By equation (12) (p. 517),

$$\Delta Q_\sigma = \Delta Q_0^- \cdot \frac{T_\sigma}{T_0} \quad \text{or} \quad \Delta Q_0^- = \Delta Q_\sigma \frac{T_0}{T_\sigma}. \quad . \quad . \quad (16)$$

It remains to show that the quantity of heat which must be supplied along the zigzag path is the same as that supplied along the actual curve. To this end we imagine a cycle performed along one of the infinitesimal triangles bordering the curve. The work done is given by the area enclosed by the triangle, and this vanishes to a higher order than the lengths of arc of the curve as the network is made finer. Since no heat is supplied along the adiabatic, the sum of the heat quantities supplied along the isothermal segment and the actual curve must be zero if all sides of the triangle are traversed in a given direction; if both segments are traversed in the same sense, the two quantities of heat must be equal.

After completing the cycle, the system under consideration and the Carnot engine are back in their original configuration, and summation over all heat exchanges gives

$$\Sigma \frac{\Delta Q_\sigma}{T_\sigma} = \frac{Q_0^-}{T_0} \quad \text{or} \quad Q_0^- = T_0 \Sigma \frac{\Delta Q_\sigma}{T_\sigma}. \quad . \quad . \quad (17)$$

In the limit, when the segments are very small, the sum is replaced by the integral

$$\oint \frac{dQ_\sigma}{T_\sigma},$$

so that for a closed path the relationship

$$\oint \frac{dQ_\sigma}{T_\sigma} = \frac{Q_0^-}{T_0} \quad \text{or} \quad Q_0^- = T_0 \oint \frac{dQ_\sigma}{T_\sigma} \quad . \quad . \quad (18)$$

is valid. If  $Q_0^-$  is negative, this means that heat is being supplied to the boiler at the expense of mechanical energy. This is entirely possible. On the other hand, a positive value of  $Q_0^-$  would mean that heat is taken from the boiler and converted into mechanical energy either in the Carnot engine or in the system under consideration. But this would be a *perpetuum mobile* of the second kind. In the most favourable case, i.e. for complete reversibility,  $Q_0^-$  is zero. In this instance, the Carnot machine uses just as much work as was done by the system in its cycle. Hence, for a reversible process between two states, the integral

$$\int_1^2 \frac{dQ_\sigma}{T_\sigma}$$

must be independent of the path, for if two curves connecting the states 1 and 2 are joined to form a closed cycle, the integral must vanish. It follows that, when the initial state 1 is assigned, the integral defines a quantity which is a function only at the final state 2. The integral

$$\int \frac{dQ_{\sigma(\text{rev})}}{T},$$

reckoned from an undetermined initial point, is called the *entropy*  $S$  of the system:

$$S = \int \frac{dQ_{\sigma(\text{rev})}}{T}. \quad . . . . . (19)$$

Thus, for a reversible passage from state 1 to state 2,

$$S_2 - S_1 = \int_1^2 \frac{dQ_{\sigma(\text{rev})}}{T}. \quad . . . . . (20)$$

For an irreversible change the sign of equality ( $=$ ) is no longer appropriate. Whether it is to be replaced by  $>$  or by  $<$  is determined by supplementing the process by a reversible one in the opposite direction, the whole process being cyclic. Since the cycle contains irreversible steps we must have, according to the statements immediately following equation (18),

$$\oint \frac{dQ_{\sigma}}{T} = \int_1^2 \frac{dQ_{\sigma(\text{irr})}}{T} + \int_2^1 \frac{dQ_{\sigma(\text{rev})}}{T} = \int_1^2 \frac{dQ_{\sigma(\text{irr})}}{T} + S_1 - S_2 < 0 \quad (21)$$

or 
$$S_2 - S_1 - \int_1^2 \frac{dQ_{\sigma(\text{irr})}}{T} > 0. \quad . . . . . (21')$$

The positive difference

$$S_2 - S_1 - \int_1^2 \frac{dQ_{\sigma}}{T}$$

is then a measure of the irreversibility of a given process. It is easy to see, in special cases, why the value of the integral

$$\oint \frac{dQ_{\sigma}}{T}$$

decreases for irreversible cyclic processes: Suppose we supply heat to the system by means of a Carnot engine, but, on account of friction, only a small part is converted into mechanical energy, or heat due to friction is generated in the system itself. Then, in order to bring the system back to the initial configuration, more heat must be withdrawn from the system than in the case of a reversible cycle.

For isolated systems, in which there is no exchange of heat with the environment, the integral

$$\int \frac{dQ}{T}$$

is zero; thus, in any change,

$$S_2 - S_1 > 0.$$

*All changes within an isolated system proceed in such way that the entropy increases; hence changes will occur in such a system until the entropy has reached its maximum value.*

Thoughtless application of  $dS = dQ/T$  to situations involving no heat exchange as such may erroneously lead to the conclusion that  $dS = 0$ . Let us illustrate what is meant, using the example of the diffusion of two gases mentioned earlier. Imagine equal volumes of the two gases, one above the other. When the separating partition is removed they will diffuse through each other and nothing else will happen. This process, occurring spontaneously, is irreversible. In a less simple way, but reversibly, the same final state can be reached by making use of the fact that cer-

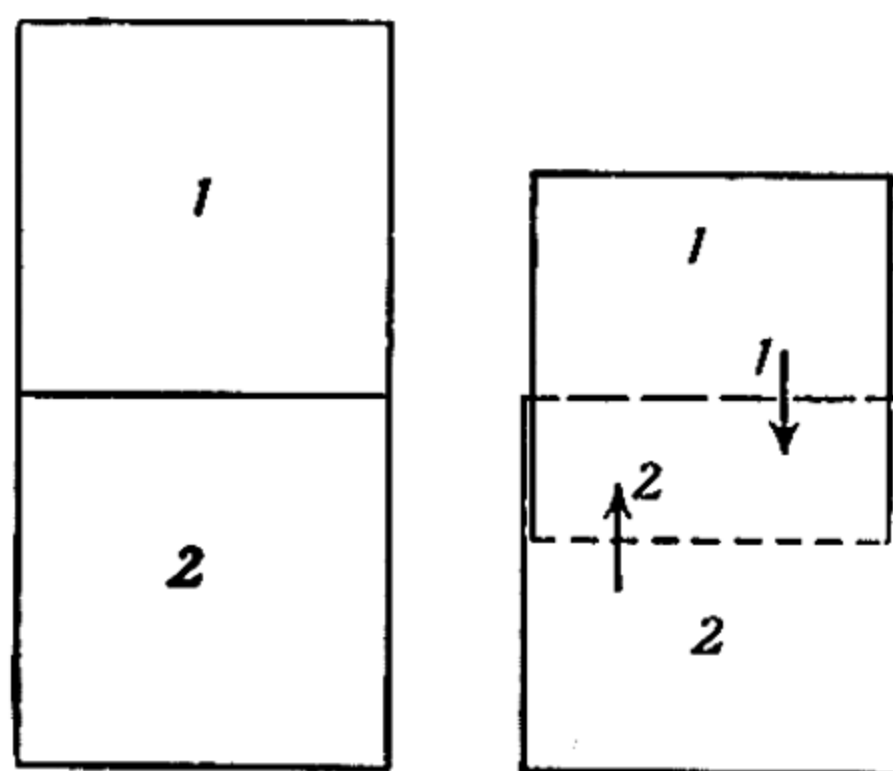


Fig. 3

tain membranes are semi-permeable. Rubber, for example, is readily permeable by carbon dioxide but not by hydrogen. For a gas that is readily transmitted by the membrane, the latter simply does not exist. Let the two gases now be placed in vessels which telescope one into another. The lower vessel has a lid that is permeable by the gas in the upper vessel, the upper one has a bottom that freely transmits the gas present in the lower vessel (fig. 3). The entire apparatus is in a vacuum. Thus no work is required when the vessels are moved together. When the two containers have been fully telescoped the gases are mixed, but the volume is half that occupied when they were mixed by irreversible diffusion. If we wish to attain this condition reversibly we must allow the mixture to expand to double its volume, in which process it does external work, but we must supply the equivalent amount of heat to the mixture in order to keep the temperature constant. This quantity of heat, divided by  $T$ , gives the change in entropy in the process of irreversible mixing. For the state attained by the reversible process, we return to initial conditions in this way: Using the work done in the expansion we compress the mixture, return the resultant heat to the heat reservoir and separate



the gases by means of the selectively permeable membranes without any expenditure of work.

Entropy is a quantity which, like temperature, may be specified numerically, usually in calories per degree. The practice of giving numerical values for this quantity is less common because earlier there was no fixed zero point for its reckoning. One thus avoided such statements as "The body has a temperature of  $100^\circ$  and an entropy of 300 cal./deg." However, since the formulation of the Nernst heat theorem (p. 556), a definite reference level for entropy is indicated and such specifications are quite in order.

*Ex.* 122. Calculate the change in entropy occurring when 10 gm. of water at  $100^\circ$  C. are mixed with 20 gm. of water at  $15^\circ$  C.

*Ex.* 123. In how many points can an adiabatic intersect an isothermal?

*Ex.* 124. What determines the position of the line  $BC$  in fig. 1 (p. 498)?

### 3. Conditions of Equilibrium for Systems under various Conditions. Thermodynamic Potentials.

Let us write (21') in differential form and express  $dQ$  in terms of the First Law, which, of course, holds whether we are dealing with reversible or irreversible processes:

$$dS - \frac{dU + \delta W^-}{T} \geq 0, \quad . . . . . (21'')$$

where  $W^-$  represents the external work, and the sign of equality applies when there is reversible conduction of heat. The notation  $\delta W^-$  indicates that the external work need not be a perfect differential of the thermodynamic variables. In the reversible case, for mechanical work,  $\delta W^-$  amounts to  $+pdV$ , where  $p$  is the equilibrium pressure—the pressure on the system at any instant. This furnishes us with a *definition of entropy in terms of the thermodynamic variables* in the form of a differential equation:

$$dS = \frac{dU + pdV}{T}. \quad . . . . . (22)$$

In order to compute the state of equilibrium that a system will assume under given external conditions, we proceed as in mechanics by using the principle of virtual displacements. For a closed system characterized by the auxiliary conditions  $U = \text{const.}$  and  $V = \text{const.}$ ,  $\delta W^-$  is also zero, and we have

$$\delta S \geq 0.$$

with the auxiliary conditions

$$\delta U = 0 \quad \text{and} \quad \delta V = 0.$$

This merely repeats the proposition, given above, that an isolated system in equilibrium assumes a condition of maximum entropy.

In general, however, the systems we deal with are not closed. Most

laboratory processes, for example, proceed at constant temperature because of heat exchange with the surroundings. For  $T = \text{const.}$  and a reversible process, (21'') yields

$$dW^- = -d(U - TS) = -dF, \quad . . . (23)$$

The function  $F$  is called the *free energy*. At constant temperature its decrease gives the maximum obtainable external work. Thus the free energy has a very vivid interpretation: Of the entire energy content, a part is available for any transformation, while the remainder is fixed and is converted into heat. In general, according to (21''),

$$\delta W^- \leq -dF. \quad . . . . . (24)$$

Now if the volume does not change (isothermal-isometric system) no work is done, and on account of  $\delta W^- = 0$ ,

$$0 \leq -\delta F$$

$$\left. \begin{array}{l} \text{or} \\ \text{with the auxiliary conditions} \end{array} \right\} \begin{array}{l} \delta F \leq 0, \\ \delta T = 0, \delta V = 0. \end{array} \quad . . . . . (25)$$

At equilibrium, an isothermal-isometric system has its minimum value of the free energy. Equilibrium is characterized by the sign of equality, because changes will take place spontaneously until this minimum value is reached and—like any extremum—this is characterized by  $\delta F = 0$ .

If the pressure in an isothermal system is held constant, then for a reversible process

$$\delta W^- = p dV = -d(U - TS),$$

or since  $p = \text{const.}$ ,

$$d(U - TS + pV) = dG = 0. \quad . . . . . (26)$$

The function  $G$  introduced here is called the *free enthalpy*, sometimes the *Gibbs potential*. It differs from the free energy by the circumstance that the enthalpy appears in place of the energy. According to (24), in any isothermal system

$$\delta W^- \leq \delta(TS - U),$$

but since  $\delta W^- \leq \delta(pV)$  and  $p = \text{const.}$ , we have

$$\delta(pV) \leq \delta(TS - U),$$

$$\left. \begin{array}{l} \text{or} \\ \text{with auxiliary conditions} \end{array} \right\} \begin{array}{l} \delta(U - TS + pV) = \delta G \leq 0, \\ \delta T = 0, \delta p = 0. \end{array} \quad . . . . . (27)$$

An isothermal-isobaric tends towards a minimum value of the free enthalpy, and this value characterizes the equilibrium.

The functions introduced here are called *thermodynamical potentials*. The name comes from their formal analogy with the potential in mechanics. In the same way as the components of force are obtained by differentiation of the mechanical potential with respect to the co-ordinates, so all other thermodynamic variables are obtained by differentiation of the thermodynamic potentials. This makes unnecessary the introduction of new constants of integration which might give rise to difficulties. The potentials must be given as functions of two properly selected independent variables. The appropriate variables are those which, together with the variation of the potential in question, occur in the auxiliary conditions. The variables appropriate to each potential are as follows:

Potential*	Independent Variables	} . . . (28)
Entropy $S$	$U, V$	
Free Energy $F$	$T, V$	
Gibbs Function $G$	$T, p$	

We then find the following relationships: From  $S = S(U, V)$  it follows that

$$dS = \left(\frac{\partial S}{\partial U}\right)_v dU + \left(\frac{\partial S}{\partial V}\right)_v dV = \frac{dU + p dV}{T}.$$

Comparison of coefficients gives

$$\frac{1}{T} = \left(\frac{\partial S}{\partial U}\right)_v; \quad \frac{p}{T} = \left(\frac{\partial S}{\partial V}\right)_v. \quad \dots (29)$$

In the same way, it follows from  $F = F(T, V)$  that

$$dF = \left(\frac{\partial F}{\partial T}\right)_v dT + \left(\frac{\partial F}{\partial V}\right)_v dV = dU - T dS - S dT,$$

which, by equation (22) (p. 524) is equal to

$$dU - dU - p dV - S dT.$$

Comparing coefficients,

$$S = - \left(\frac{\partial F}{\partial T}\right)_v, \quad p = - \left(\frac{\partial F}{\partial V}\right)_v. \quad \dots (30)$$

\* The notation used here for the potentials is largely that of Clausius. Gibbs called these functions  $\eta$ ,  $\psi$ , and  $\zeta$  in the above order. He also denoted the internal energy by  $\epsilon$  and the absolute thermodynamic temperature by  $t$ . See J. W. Gibbs, *Collected Works*, 1, New York, Longmans, Green & Co. (1928).



Also, from  $G(T, p)$  we find

$$dG = \left(\frac{\partial G}{\partial T}\right)_p dT + \left(\frac{\partial G}{\partial p}\right)_T dp = dU - T dS - S dT + p dV + V dp,$$

which is equal to

$$-S dT + V dp$$

by equation (22). Hence

$$S = -\left(\frac{\partial G}{\partial T}\right)_p; \quad V = \left(\frac{\partial G}{\partial p}\right)_T. \quad \dots \quad (31)$$

The relationships derived from  $H = H(S, p)$  are less important. It follows from equations (31) and (32) that the free energy and the Gibbs Potential satisfy the differential equations

$$F = U + T \left(\frac{\partial F}{\partial T}\right)_p, \quad \dots \quad (32)$$

$$G = U + pV + T \left(\frac{\partial G}{\partial T}\right)_p. \quad \dots \quad (33)$$

These are obtained by substituting the expressions for  $S$  in the equations defining the functions  $F$  and  $G$ .

All expressions entering into these thermodynamic potentials are such that they merely add when two systems are placed in contact and allowed to come to equilibrium. For this reason the potentials themselves are additive quantities. In particular, if the mass of a system is multiplied by  $n$ , the values of the potentials increase by this same factor.

#### 4. Connexion between the Internal Energy and the Equation of State.

The Second Law is capable of giving a purely thermodynamic derivation of the dependence of the energy  $u$  on  $T$  and  $v$ —a connexion which was obtained in the previous chapter (p. 503) from partly kinetic considerations. If we take  $T$  and  $v$  as independent variables, we have for one mol:

$$\begin{aligned} ds &= \left(\frac{\partial s}{\partial T}\right)_v dT + \left(\frac{\partial s}{\partial v}\right)_T dv \\ &= \frac{1}{T} \left[ \left(\frac{\partial u}{\partial T}\right)_v dT + \left(\frac{\partial u}{\partial v}\right)_T dv + p dv \right]. \quad \dots \quad (34) \end{aligned}$$

Comparing coefficients,

$$\left(\frac{\partial s}{\partial T}\right)_v = \frac{1}{T} \left(\frac{\partial u}{\partial T}\right)_v, \quad \dots \quad (35)$$

$$\left(\frac{\partial s}{\partial v}\right)_T = \frac{1}{T} \left[ \left(\frac{\partial u}{\partial v}\right)_T + p \right]. \quad \dots \quad (36)$$

If we differentiate (35) with respect to  $v$  and differentiate (36) with respect to  $T$ , comparison gives

$$\left(\frac{\partial u}{\partial v}\right)_T = T \left(\frac{\partial p}{\partial T}\right)_v - p = T^2 \left[\frac{\partial(p/T)}{\partial T}\right]_v. \quad (37)$$

Using the equation of state of an ideal gas, this relationship leads to the independence of the internal energy and the volume; if the van der Waals equation is used, we obtain equation (18) (p. 505). From equation (11) (p. 504), we obtain the general relationship for the difference of the specific heats:

$$c_p - c_v = T \left(\frac{\partial p}{\partial T}\right)_v \left(\frac{\partial v}{\partial T}\right)_p. \quad (38)$$

*Ex. 125.* In a voltaic cell whose electromotive force is  $\Phi$  an amount of work  $\Phi de$  is done when a quantity of electricity  $de$  passes. The expression  $\Phi de$  thus appears in place of the mechanical work  $p dv$ . The energy change is determined by the fact that an amount of heat equal to the heat of reaction per gramme equivalent is generated when 96,496 coulombs of charge have been transported. What relationship exists between the heat of reaction, the electromotive force, and its temperature coefficient? *Hint:* Introduce the charge  $e$  and the temperature  $T$  as independent variables.

*Ex. 126.* Show that for  $T$  and  $p$  as independent variables the enthalpy  $h$  takes the place of  $u$  and that for  $(2h/\partial p)_T$  we get a formula analogous to (37).

## 5. Electrocaloric and Magnetocaloric Phenomena.

In our considerations of energy relating to electrical phenomena we did not take heat energy into account. Strictly speaking, we would be concerned with the quantity designated there by  $u$ —the *free* energy; but since free energy and true energy differ but little at room temperature (they are identical at absolute zero), little error is involved. It is only when we study the thermal effects connected with polarization that this difference must be kept in mind.

As a starting point we use equation (36'), p. 290, according to which an amount of energy  $\mathbf{E} d\mathbf{P}$  is supplied to each cubic centimetre of a body when its polarization is changed. If, in addition, heat is supplied (but not mechanical energy), then the First Law prescribes that

$$du = \delta q + \mathbf{E} d\mathbf{P}. \quad (39)$$

We are able to work with merely the magnitudes of the vectors, since  $\mathbf{P}$  and  $\mathbf{E}$  have the same direction. Comparing this formula with the one that holds for mechanical work, we see that  $-E$  appears in place of  $p$  and  $P$  replaces  $v$ . Inasmuch as the polarization  $P$  usually refers to unit volume, the formula as it stands is valid for a single cubic centimetre of material. However, it may be thought of as applying to one mol if we interpret  $P$  to be the polarization per mol. The same equation is valid for the magnetic case also. In order to avoid confusion with

other quantities, let the magnetization per mol be denoted by  $l$ . Since the magnetic case is of much greater practical importance, the further development will be given in terms of these phenomena. If the mechanical quantities in the equations of the last section are replaced by the magnetic ones, we obtain at once the difference of the specific heats at constant magnetization and at constant field. To interpret the result, consider a paramagnetic substance placed between the poles of a magnet and warmed. This lowers its paramagnetic susceptibility and so, at constant field strength, its magnetization. Instead, however, we could hold the magnetization constant by increasing the field. The detailed computation is not difficult (cf. *Ex.* 127, p. 530).

Of considerable experimental importance in recent years is the temperature change accompanying adiabatic processes, because by their use we are able to approach closer to the absolute zero than by any other means. Since adiabatic processes are characterized by constant entropy, we must write down this function for a magnetized body. Transcription of equation (37) yields

$$\left(\frac{\partial u}{\partial l}\right)_T = -T^2 \left[ \frac{\partial(H/T)}{\partial T} \right]_l. \quad \cdot \cdot \cdot \cdot (40)$$

In studying paramagnetic susceptibility we found that as long as the elementary magnets do not influence each other—and hence the Curie law rather than the Weiss law is valid— $l$  is a function only of  $H/T$ . Reciprocally,  $H/T$  is then a function of  $l$  alone and so the derivative on the right side of the above equation vanishes. Thus *the internal energy of a magnetized body is independent of the magnetization*, just as the internal energy of an ideal gas is independent of the volume. We conclude that a material obeying the Curie law corresponds to an ideal gas. As in the case of gases, the departures from the law, which manifest themselves in an internal field, become increasingly important at low temperatures. For an ideal magnetic material, the transcription of (34) becomes

$$ds = \frac{1}{T} \left[ \left( \frac{\partial u}{\partial T} \right)_l dT - H dl \right]. \quad \cdot \cdot \cdot \cdot (41)$$

The second term in brackets represents the magnetic contribution to the entropy. The specific heat is to be substituted for the entire first term; and since we deal with solids or liquids no distinction need be made between  $c_p$  and  $c_v$ . Note that, according to the above, there is no influence of the magnetization on the internal energy and hence on the specific heat.

In order to evaluate the magnetic share of the entropy we start with the fact that the magnetization  $l$  of a paramagnetic body is a function of  $\mu H/kT$ , where  $\mu$  is the magnetic moment of the atom in



question. If this fraction is introduced as a variable  $x$ , the magnetic contribution to the entropy becomes, by integration of the second part of (41),

$$s_m = - \int \frac{H}{T} dl = - \frac{k}{\mu} \int \frac{\mu H}{kT} dl = - \frac{k}{\mu} \int x f'(x) dx \\ = - \frac{k}{\mu} \left[ x f(x) - \int f(x) dx \right]. \quad (42)$$

Putting in the function  $\mu L \tanh x$  which holds in very many cases and noting that  $kL = R$ , we have

$$S_m = -R(x \tanh x - \log \cosh x) = \Phi \left( \frac{\mu H}{kT} \right). \quad (43)$$

In an adiabatic process the total entropy is constant, so that

$$s = \int \frac{c}{T} dT + \Phi \left( \frac{\mu H}{kT} \right) = \text{const.} \quad (44)$$

On account of the dominance of the first term the function  $\Phi$  is negative. If the field is switched on at a given temperature  $T$  the negative term appears and so the first term must increase in order that the entropy remain constant. This means that the temperature must increase to a value  $T + \Delta T$  given by

$$\int_T^{T+\Delta T} \frac{c}{T} dT = - \Phi \left( \frac{\mu H}{kT} \right). \quad (45)$$

For the small term  $\Phi(\mu H/kT)$  we may use merely the initial temperature  $T$ . Assuming an average value  $\bar{c}$  of the specific heat over the interval  $\Delta T$ , (45) yields

$$\frac{\Delta T}{T} = - \frac{1}{\bar{c}} \Phi \left( \frac{\mu H}{kT} \right). \quad (45')$$

Conversely, on suddenly removing the field there must occur a drop in temperature whose magnitude is given by this same equation; and in the neighbourhood of absolute zero, where other adiabatic processes fail, this temperature drop may be of considerable importance, granted only that  $l$  depends solely on  $H/T$  (see above). By this means, temperatures of only a few hundredths of a degree absolute have been attained. Since other methods will not work, the measurement of these temperatures is accomplished by means of susceptibility measurements. Here the problem arises of computing the true temperatures from readings on the Curie scale, the two differing because of the appearance of ferromagnetism. The difficulty has been overcome for a substance such as ferric ammonium alum.

*Ex. 127.* Compute the difference of the specific heats at constant field and constant magnetization in a region where  $l = CH/T$ .

## CHAPTER XXXII

### APPLICATION OF THE SECOND LAW TO THE CALCULATION OF THE EQUILIBRIUM OF THERMODYNAMIC SYSTEMS

#### 1. Gibbs' Phase Rule: a General Theorem on the Maximum Number of possible Phases.

Consider a perfectly general system containing  $\alpha$  independent components (substances) in  $\beta$  phases. We shall find presently that  $\beta$  has a maximum value, determined by the number of independent components. We indicate the components by subscripts and the phases by superscripts attached to the letters. Thus the mass of the  $i$ th component in the  $k$ th phase is represented by  $M_i^k$ .

The equilibrium is considered when the system is at a given temperature and a given pressure; thus we consider it an isothermal-isobaric system rather than an isolated one. In this case the Gibbs Potential

$$G = U - TS + pV \quad . \quad . \quad . \quad . \quad . \quad (1)$$

is a minimum, according to p. 525. Of the quantities entering here,  $T$  and  $p$  are entirely independent of the mass of the system. On the other hand,  $U$ ,  $V$  and  $S$  are proportional to the mass, for given temperature and pressure. From this, one sees that  $G$  is also proportional to the mass; thus if all masses are doubled,  $G$  becomes twice as great. In general

$$G(qM_1, qM_2, \dots, qM_\alpha) = qG(M_1, M_2, \dots, M_\alpha). \quad (2)$$

But this is the definition of a homogeneous function of first degree in the masses. For homogeneous functions of  $n$ th degree the Euler theorem

$$M_1 \frac{\partial G}{\partial M_1} + M_2 \frac{\partial G}{\partial M_2} + \dots + M_\alpha \frac{\partial G}{\partial M_\alpha} = nG(M_1, M_2, \dots, M_\alpha) \quad (3)$$

is valid. In our case this is

$$M_1 \frac{\partial G}{\partial M_1} + M_2 \frac{\partial G}{\partial M_2} + \dots + M_\alpha \frac{\partial G}{\partial M_\alpha} = G(M_1, M_2, \dots, M_\alpha). \quad (4)$$

To show this, it is necessary only to differentiate equation (2) with

respect to  $q$  and set  $q = 1$ . The right member is a homogeneous function of first degree in the masses. In order that the left member be such a function also, the derivatives  $\partial G/\partial M_i$  must be homogeneous functions of degree zero, i.e. *the derivatives may be functions only of the ratios of the masses*.

After these preliminary remarks concerning the properties of the function  $G$ , we return to the question under consideration. The quantities  $U$ ,  $V$  and  $S$  are additive, e.g. the internal energy of a system is equal to the sum of the energies of its parts. Hence if we distinguish the Gibbs Potentials of the several phases by accents, we have for the entire system

$$G = G' + G'' + \dots G^\beta. \quad . . . . . (5)$$

The  $G^k$  have, of course, the properties deduced above. Besides, they depend also on  $T$  and  $p$ . The condition of equilibrium is

$$\delta G = \delta G' + \delta G'' + \dots \delta G^\beta = 0, \quad . . . . . (6)$$

with the supplementary conditions that  $\delta T = 0$  and  $\delta p = 0$ . In addition to this, the mass of each component must remain unaltered; the distribution of a given substance among the various phases may change, but the total mass of the substance remains constant. We then have the following  $\alpha$  auxiliary conditions also:

$$\left. \begin{aligned} \delta M_1' + \delta M_1'' + \dots \delta M_1^\beta &= 0, \\ \delta M_2' + \delta M_2'' + \dots \delta M_2^\beta &= 0, \\ . & . . . . . \\ \delta M_\alpha' + \delta M_\alpha'' + \dots \delta M_\alpha^\beta &= 0. \end{aligned} \right\} . . . . . (7)$$

Taking the variation of  $G$  while  $T$  and  $p$  are held constant, we obtain the equation

$$\begin{aligned} \delta G = 0 &= \frac{\partial G'}{\partial M_1'} \delta M_1' + \frac{\partial G'}{\partial M_2'} \delta M_2' + \dots \frac{\partial G'}{\partial M_\alpha'} \delta M_\alpha' \\ &+ \frac{\partial G''}{\partial M_1''} \delta M_1'' + \frac{\partial G''}{\partial M_2''} \delta M_2'' + \dots \frac{\partial G''}{\partial M_\alpha''} \delta M_\alpha'' \\ &+ \dots + \frac{\partial G^\beta}{\partial M_1^\beta} \delta M_1^\beta + \frac{\partial G^\beta}{\partial M_2^\beta} \delta M_2^\beta + \dots \frac{\partial G^\beta}{\partial M_\alpha^\beta} \delta M_\alpha^\beta. * \end{aligned} \quad (8)$$

Since the variations are not arbitrary, but are subject to the conditions (7), we apply the method of Lagrangian Multipliers (cf. p. 115), multiply (7) by  $\lambda_1, \lambda_2, \dots$ , and add them to the principal

\* The partial derivatives occurring here are sometimes referred to collectively as the Gibbs Potentials. Gibbs pointed out that they furnish a quantitative measure of the concept of chemical affinity. See p. 554.



equation (8). Since the coefficients of the  $\delta M_i^k$  must vanish separately, we obtain the system of equations

$$\left. \begin{aligned} \frac{\partial G'}{\partial M_1'} + \lambda_1 &= 0, & \frac{\partial G'}{\partial M_2'} + \lambda_2 &= 0, & \dots & \frac{\partial G'}{\partial M_a'} + \lambda_a = 0, \\ \frac{\partial G''}{\partial M_1''} + \lambda_1 &= 0, & \frac{\partial G''}{\partial M_2''} + \lambda_2 &= 0, & \dots & \frac{\partial G''}{\partial M_a''} + \lambda_a = 0, \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \frac{\partial G^\beta}{\partial M_1^\beta} + \lambda_1 &= 0, & \frac{\partial G^\beta}{\partial M_2^\beta} + \lambda_2 &= 0, & & \frac{\partial G^\beta}{\partial M_a^\beta} + \lambda_a = 0. \end{aligned} \right\} \quad (9)$$

These equations,  $\alpha\beta$  in number, must be satisfied. How many variables are at our disposal? First there are the  $\alpha$  multipliers—then the derivatives depend on  $T$ ,  $p$  and the  $(\alpha - 1)$  mass ratios, i.e. on  $(\alpha + 1)$  variables. Inasmuch as  $T$  and  $p$  are the same for every phase, we have in all  $\alpha + \beta (\alpha - 1) + 2$  free variables. Since the number of equations can at most equal the number of variables,

$$\alpha\beta \leq \alpha + \beta(\alpha - 1) + 2$$

**or**

$$\beta \leq \alpha + 2. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (10)$$

The number of phases may be, at most, two greater than the number of independent components. When  $\beta$  is equal to  $\alpha + 2$ ,  $T$  and  $p$  are completely determined, since solution of the system of equations yields definite numerical values for these quantities. Thus a system consisting of a single substance can have three phases at most. For example, at a definite temperature and pressure water, water vapour, and ice can co-exist in thermodynamic equilibrium, but a second form of ice cannot be included. In the latter instance there could be no equilibrium, changes would take place until one phase would be entirely consumed.

If there are but two phases we still have one variable at our disposal. In a system consisting of water and water vapour the temperature, for example, is at our disposal, but the pressure is uniquely determined by the temperature and is not a second independent variable as in the case of a single phase system—say a gas. In general, the difference between the highest possible number of phases and the number of phases actually present is the number of degrees of freedom, *f*. Thus

$$f = \alpha + 2 - \beta. \quad . \quad . \quad . \quad . \quad . \quad . \quad (11)$$

As an example of a system consisting of two substances, consider a solution of sal ammoniac in water. Assume a very dilute solution with a second phase consisting of the water vapour above the solution. If the solution is cooled, ice begins to form at a temperature a little

below  $0^{\circ}\text{C}$ . We now have a third phase: ice. According to the phase rule there is still one variable free; take this to be the temperature. The concentration is determined by this variable, for if we continue to cool the system, more ice is formed, the solution becomes more concentrated and so its freezing-point sinks. The last fact is one of experience; a thermodynamic derivation of the result will be given on p. 545. Finally, the solution becomes saturated and the solid salt also precipitates from solution. The temperature at which this occurs is called the *eutectic point*; this is the lowest temperature the system can attain. At this quadruple point the vapour, solution, ice and salt are all in equilibrium. If, on the other hand, we start with a highly concentrated solution, the solid salt comes out of solution as the third phase when the temperature is lowered. Additional cooling causes the concentration to decrease further. Thus the curve of concentration *versus* temperature is also the solubility curve. If the eutectic point is reached from this side, there is again a simultaneous precipitation of ice and salt. A solution having the concentration corresponding to the eutectic point has the lowest possible freezing-point.

## 2. The Vapour Pressure Curve and the Melting-point Curve.

The method of solving thermodynamic problems which was used in the preceding section consists in seeking the extremal value of the function  $S$  or  $F$  or  $G$ , whichever is pertinent to the situation at

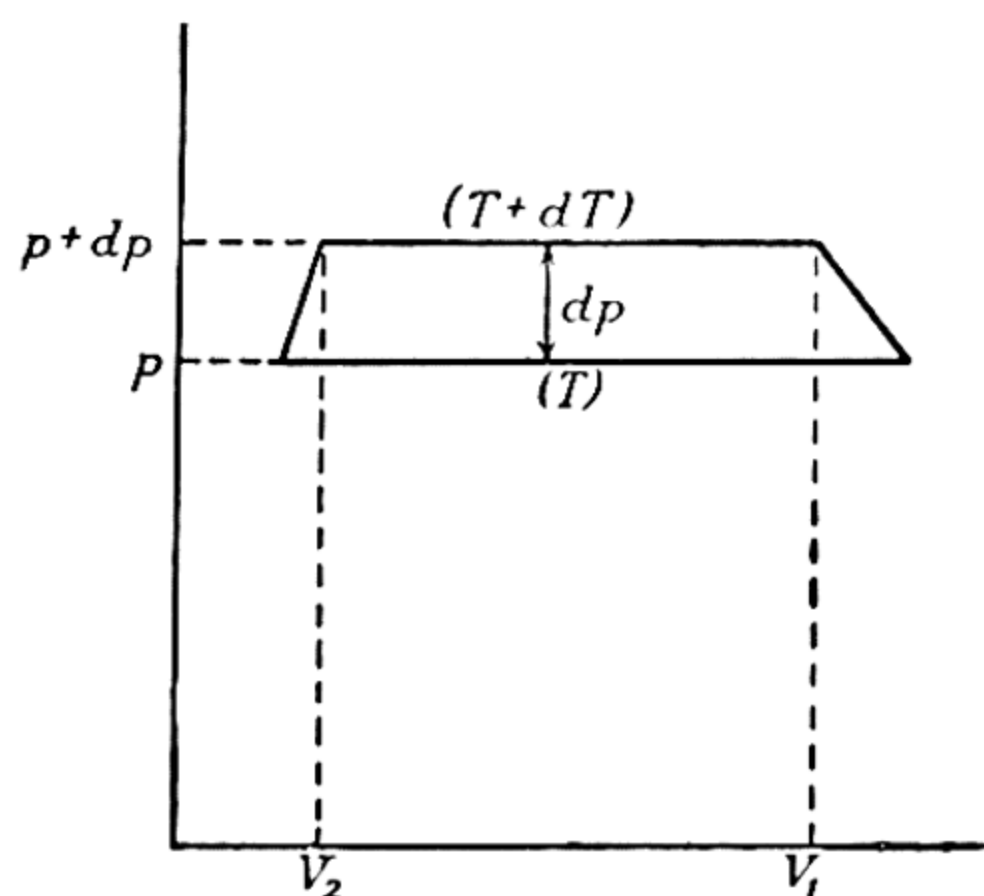


Fig. 1

hand. The procedure demands little thought but does require some computational work. It is, however, possible to proceed otherwise by considering a cyclic process for each special case, and applying to it the two thermodynamic laws. We shall compare the two methods in deducing the vapour pressure curve. Let us formulate the problem in this way: Consider a quantity of liquid in a closed vessel together with its vapour. The vapour pressure is measured by means of a mano-

meter. Since we are dealing with a two-phase system involving only one substance, there will be but a single degree of freedom, which we take to be the temperature. The volume plays no part, for if we decrease the volume the pressure does not increase—as it would if a gas alone were present. Instead, some of the vapour condenses. Using the device of a cyclic process for determining the relation of  $p$  and  $T$ , let



us consider the cycle represented by fig. 1. At a temperature  $T + dT$  we vaporize one mol of liquid, thus moving a piston forward against the pressure  $p + dp$ . Next, we relieve the pressure somewhat by decreasing the application of heat and arrive at the temperature  $T$ . At this temperature we reduce the volume until the vapour has again just condensed, after which we apply a small amount of heat to bring the substance back to the temperature  $T + dT$ . Neglecting all second-order quantities, as well as the minute quantities of heat involved in passing between the two neighbouring temperatures, the work obtained turns out to be  $dp(v_1 - v_2)$ , where the subscript 1 refers to the gas and 2 to the liquid. In the evaporation process an amount of heat equal to the heat of vaporization  $q_{T+dT} \approx q_T$  was communicated. By the Second Law, a fraction  $dT/T$  of this was converted into work, so that

$$dp(v_1 - v_2) = \frac{q_T dT}{T}, \quad \text{or} \quad \frac{dp}{dT} = \frac{q_T}{(v_1 - v_2)T}. \quad (12)$$

This *Clausius-Clapeyron differential equation of the vapour pressure curve* is rigorously true, but before performing the integration we make some simplifying assumptions. For one thing, we neglect the volume  $v_2$  of the liquid in comparison with that of the vapour; also, we assume that the equation of state of an ideal gas applies to the vapour; finally, we work with constant specific heats. First consider the liquid to evaporate at  $T = 0$  and  $p = p$ , and the resulting vapour to be warmed to  $T^\circ$ . Alternatively, let the liquid first be warmed from  $0^\circ$  to  $T^\circ$  and then vaporized. We have

$$q_0 + c_p T = q_T + c_2 T \quad \text{or} \quad q_T = q_0 + (c_p - c_2)T, \quad (13)$$

where  $q_0$  is the heat of vaporization at  $T = 0$ . Then

$$\frac{dp}{dT} = \frac{[q_0 + (c_p - c_2)T]_p}{RT^2}. \quad (14)$$

Integrating,

$$\log p = -\frac{q_0}{RT} + \frac{c_p - c_2}{R} \log T + \log A; \quad (15)$$

$$\text{or} \quad p = A e^{-\frac{q_0}{RT}} T^{\frac{c_p - c_2}{R}}. \quad (16)$$

Since the influence of the last factor is small, the equation shows that the vapour pressure rises rapidly—almost exponentially—with the temperature.

Let us now deduce the vapour pressure curve by the other method as well. In order not to repeat exactly what was done, let us take account here of the temperature dependence of the specific heats. At equilibrium the temperature is, of course, the same throughout the liquid-vapour system. Inasmuch as we seek the equilibrium position for a given value of  $T$  and of  $p$ , we must find the minimum value of



$G$ . If there are  $n_1$  mols of vapour and  $n_2$  mols of liquid, the total free enthalpy is

$$G = n_1 g_1 + n_2 g_2, \quad . \quad . \quad . \quad . \quad . \quad (17)$$

where the  $g_i$  refer to one mol. The variation is to be taken under the conditions

$$T = \text{const.}, \quad p = \text{const.}, \quad n_1 + n_2 = \text{const.} \quad . \quad . \quad (18)$$

Since the  $g_i$  depend only upon  $T$  and  $p$ , they must remain constant, on account of the auxiliary conditions. We have thus

$$0 = \delta n_1 g_1 + \delta n_2 g_2. \quad . \quad . \quad . \quad . \quad . \quad (19)$$

We also have

$$0 = \delta n_1 + \delta n_2, \quad . \quad . \quad . \quad . \quad . \quad (20)$$

The method of undetermined multipliers yields

$$g_1 + \lambda = 0, \quad g_2 + \lambda = 0,$$

$$\text{or} \quad g_1 - g_2 = u_1 - u_2 + p(v_1 - v_2) + T(s_2 - s_1) = 0. \quad (21)$$

We must now try to compute the separate terms of this equation. For this purpose we treat the vapour as an ideal gas. In the neighbourhood of the point of condensation this may involve considerable error. We also neglect the volume of the liquid in comparison with that of the vapour in calculating the factor  $(v_1 - v_2)$ . Under these conditions, by (13), p. 504,

$$\left. \begin{aligned} u_1 &= u_{1,0} + \int_0^T c_{v1} dT, \\ u_2 &= u_{2,0} + \int_0^T c_{v2} dT.^* \end{aligned} \right\} \quad . \quad . \quad . \quad . \quad (22)$$

We have, in addition,

$$pv_1 = RT = \int_0^T R dT. \quad . \quad . \quad . \quad . \quad (23)$$

The expression  $(u_{1,0} - u_{2,0})$  represents the value at absolute zero of the difference in energy between vapour and liquid, i.e. the energy which must be supplied to the liquid to overcome the internal forces of attraction. This is then  $q_0$ , the value at absolute zero of the work of vaporization (or since it is usually given in calories, the *heat of vaporization*). Using equation (23) and Mayer's equation [(14) p. 505], there results

$$u_1 - u_2 + p(v_1 - v_2) = q_0 + \int_0^T c_p dT - \int_0^T c_2 dT. \quad (24)$$

\* For the liquid we start also with the relationship  $(\partial u / \partial T)_v = c_v$ ; in the case of a liquid or a solid, however, the external work may be neglected, so that the two specific heats may be considered equal.

To calculate the expression  $T_s$  we turn to the differential equation for the entropy:

$$ds = \frac{1}{T} (du + p dv).$$

For the liquid we can again neglect the work done in expansion, just as we need not distinguish between the two specific heats for liquids and solids. Then we have

$$ds_2 = c_2 \frac{dT}{T}.$$

Integrating between the limits 0 and  $T$ ,

$$s_2 = \int_0^T \frac{c_2 dT}{T} + s_{2,0} \cdot \cdot \cdot \cdot \cdot \cdot \cdot \quad (25)$$

The quantity  $s_{2,0}$ —the entropy of the liquid at the absolute zero—is undetermined in the classical thermodynamics. We shall find later that for a homogeneous liquid this quantity must be zero. Expressing  $p$  in terms of  $v$  in the term  $p dv$ , we have for the vapour

$$ds_1 = \frac{c_v dT}{T} + \frac{R dv}{v}.$$

Indefinite integration gives

$$\begin{aligned} s_1 &= \int \frac{c_v dT}{T} + R \log v \\ &= \int \frac{c_v dT}{T} + R \log T + R \log R - R \log p + \text{const.} \quad . \quad (26) \end{aligned}$$

Now take the absolute zero as lower limit, and combine all terms not dependent upon  $T$  and  $p$  into a single constant  $s_{1,0}$ ; this represents the entropy of the gas at  $T = 0$  and  $p = 1$ . Since

$$\log T = \int \frac{dT}{T} \text{ and } c_p - c_v = R,$$

the entropy at temperature  $T$  is given by

$$s_1 = \int_0^T \frac{c_p dT}{T} - R \log p + s_{1,0} \quad . \quad . \quad . \quad (26')$$

Now the specific heat of an ideal gas does not vanish at  $T = 0$ , and so questions of convergence arise in connexion with the integral. In order to clarify the situation, consider the specific heat  $c_p$  to be made up of a part  $c_{p0}$  which is independent of the temperature and another part  $c_{pT}$  which, according to experience, approaches zero rapidly as  $T$  decreases. This latter part is absent altogether for a monatomic gas. We thus take

$$C_p = C_{p0} + C_{pr}, \quad \cdot \cdot \cdot \cdot \cdot \cdot (27)$$

For  $p = 1$ , we have for a very small lower limit,

$$s_1(T, 1) - s_1(T_0, 1) = \int_0^T \frac{c_{pT} dT}{T} + c_{pc} \log T - c_{pc} \log T_0. \quad (28)$$

Because of the rapid decline in the value of  $c_{pT}$ , we have written the lower limit of the integral as zero. Since the contribution of  $c_{pT}$  to the integral is negligible even at  $T_0 = 1^\circ$  absolute, we may interpret  $s_{1,0}$  approximately as the entropy at  $T = 1$  and  $p = 1$ . Strictly, it is the limiting value of the difference  $s_1(T_0, 1) - c_{pc} \log T_0$  as  $T_0$  approaches zero.\*

Replace also  $c_p$  by  $c_{pc} + c_{pT}$  and substitute (24), (25) and (26') in equation (21). Solving for  $\log p$ , we obtain the vapour pressure formula:

$$\begin{aligned} \log p = & -\frac{q_0}{RT} + \frac{1}{RT} \int_0^T c_2 dT - \frac{1}{RT} \int_0^T c_{pT} dT + \frac{1}{R} \int_0^T \frac{c_{pT} dT}{T} \\ & + \frac{c_{pc}}{R} \log T - \frac{1}{R} \int_0^T \frac{c_2 dT}{T} + \frac{s_{1,0} - s_{2,0} - c_{pc}}{R}. \end{aligned} \quad (29)$$

If we were to assume constant specific heats we would again obtain equation (15), p. 535, except that the constant of integration would be written differently. Of course, differentiation would again yield the Clausius-Clapeyron equation.

The constant  $(s_{1,0} - s_{2,0} - c_{pc})/R$  appearing in (29) is called the "chemical constant" of the gas, since it occurs also when considering chemical equilibrium. It is not determinable by the two thermodynamic laws, but may be evaluated theoretically by calculations involving quantum statistics. This constant may be determined experimentally by means of the observed vapour pressures, however, since all remaining quantities in equation (29) are sufficiently known.

It is customary here to measure the pressure in atmospheres, to replace logarithms to base  $e$  by ordinary logarithms and to measure all quantities in thermal units ( $R = 1.985$  cal). The result is

$$\begin{aligned} \log p_{\text{atm.}} = & -\frac{q_0(\text{cal})}{4.571 T} + \frac{c_{pc}}{1.985} \log T + \frac{1}{4.571 T} \int_0^T (c_2 - c_{pT}) dT \\ & + \frac{1}{4.571} \int_0^T \frac{(c_{pT} - c_2) dT}{T} + i. \quad \dots \dots \dots (30) \end{aligned}$$

As we shall find later,  $s_{2,0}$  vanishes, so that

$$i = \frac{1}{4.571} (s_{1,0} - c_{pc}) - \log (1.013 \times 10^6).$$

For a monatomic gas of atomic weight  $A$  the quantum statistics (p. 638) gives the value

$$i = -1.59 + \frac{3}{2} \log A. \quad \dots \dots \dots (31)$$

\* That this difference has a definite limit is a consequence of the empirically verified Nernst heat theorem (p. 556).



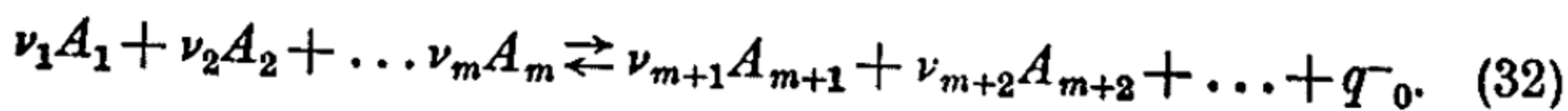
All the equations derived above are valid also for the transition from the solid state (subscript 3) to the gaseous state. We need only replace the subscript 2 by 3 throughout;  $q$  is then the *heat of sublimation*. On the other hand, the integrated formulæ contain an approximation which prevents their application to the transition between the solid and liquid states, for here the two volumes  $v_2$  and  $v_3$  are approximately equal, so that it is no longer allowable, as it was in the case of a gas, to neglect one as compared with the other. The Clausius-Clapeyron equation, however, is exact and may be applied also to the phenomenon of melting. Inverting the formula, it gives the variation of the melting-point with pressure:

$$\frac{dT}{dp} = \frac{T(v_2 - v_3)}{q_r} \quad \cdot \cdot \cdot \cdot \cdot \quad (12')$$

The quantity  $q_r$  represents the energy which must be supplied to the system to bring it from the state of higher subscript to the lower—in our case the *heat of fusion*. Since this is always positive, the sign of  $dT/dp$  depends only on that of  $v_2 - v_3$ . If, as is usually the case, the solid is denser than the liquid,  $v_2 > v_3$ , and the melting-point rises with increasing pressure. In the opposite case, e.g. for water, the melting-point falls as the pressure increases.

### 3. Chemical Equilibrium in a Mixture of Ideal Gases. The Law of Mass Action.

We now consider a system consisting of only one phase—the gaseous—but of several substances, all of which are ideal gases. The gases will react with each other chemically to a certain extent, but a chemical reaction is never complete in the sense of the equation. For example, if hydrogen and oxygen are allowed to combine to form water, there remains always a certain amount of the original substances  $H_2$  and  $O_2$ . The fraction of the original material remaining increases with the temperature. We set ourselves the task of determining the distribution in the state of equilibrium as a function of the pressure and temperature. Let the volume of the system be  $V$  and let the number of mols of the  $i$ th gas be  $n_i$ . Represent the possible reaction by the equation



In order to avoid mistakes in sign, we agree to write the equation in exothermal form, in which the heat of reaction is positive. In altering the numbers of mols, it must be remembered that they must satisfy the chemical equation, i.e. they must be in the ratio of the number of mols formed or removed. On account of the sign, we specify that the numbers of mols present before the exothermal reaction (i.e. those

on the left) are positive—those on the right, negative. Hence \*

$$\begin{aligned} \delta n_1 : \delta n_2 : \dots \delta n_m : \delta n_{m+1} : \dots \\ = + \nu_1 : + \nu_2 : \dots - \nu_{m+1} : - \nu_{m+2} : \dots \end{aligned} \quad (33)$$

Since we investigate the equilibrium at a given temperature and pressure, the variation of the function  $G = U + pV - TS$  must again vanish. The several terms may be computed as follows: The energy of the mixture is, naturally, equal to the sum of the energies of the components:

$$U = n_1 u_1 + n_2 u_2 + \dots = \sum n_i u_i. \quad . \quad . \quad . \quad (34)$$

For the volume we have, by p. 496,

$$V = \sum n_i \frac{RT}{p}. \quad . \quad . \quad . \quad . \quad . \quad (35)$$

The calculation of the entropy of the mixture as a function of  $T$  and  $p$  is somewhat more difficult. It is true that we know the entropy of a single ideal gas (§ 2, p. 537). Moreover, if we have two gases side by side, both at the same pressure, their entropy is equal to the sum of the individual entropies, but if the gases are allowed to mix, the very process of mixing is a self-maintaining one and involves an increase of entropy; this process is simply the expansion of both gases to the total volume of the vessel, whereby the partial pressure drops but the total pressure remains constant. It would seem more plausible that the entropy of the mixture should be equal to the sum of the entropies of the separate gases when each one occupies the volume of the mixture at a pressure equal to its partial pressure, for the process described on p. 523 permits the two gases to mix without change of entropy in such way that the partial pressure remains the same after mixing.

In order to simplify the calculation, we assume the specific heats to be constant and obtain, as in equation (26') (p. 537),

$$s_i = c_{pi} \log T - R \log p_i + s_{i0}, \quad . \quad . \quad . \quad (36)$$

where  $s_{i0}$  embraces all constant terms, including the term appearing for constant specific heat,  $c_p \log T_0$ , as  $T_0 \rightarrow 0$ !

Expressing the partial pressure in terms of the total pressure, we have, by p. 496,

$$\begin{aligned} s_i &= c_{pi} \log T - R \log n_i - R \log p + R \log \sum n_i + s_{i0} \\ &= s_i(T, p) - R \log n_i + R \log \sum n_i, \end{aligned} \quad . \quad . \quad . \quad (37)$$

and 
$$S = \sum n_i s_i(T, p) - R \sum n_i \log n_i + R \sum n_i \log \sum n_i. \quad . \quad (38)$$

\* With this agreement as to sign, the changes  $\delta n_i$  signify the numbers of mols formed in the *endothermal* reaction.

The entropy of the  $i$ th gas at a temperature  $T$  and pressure  $p$  is denoted by  $s_i(T, p)$ . Multiplying (35) by  $p$ , adding it to equation (34) (p. 540) and subtracting (38) multiplied by  $T$ , and making use of the fact that  $RT = pv_i$ , we obtain the Gibbs Potential of the system:

$$G = \sum n_i g_i(T, p) + RT \sum n_i \log n_i - RT \sum n_i \log \sum n_i, \quad (39)$$

where the  $g_i(T, p)$  denote the Gibbs Potentials of the separate gases at temperature  $T$  and pressure  $p$ . In order to determine the equilibrium it is necessary to take the variation of  $G$  with the auxiliary conditions (33) (p. 540), as well as with  $\delta T = \delta p = 0$ . Since the  $g_i$  depend only upon  $T$  and  $p$  their variations vanish and we obtain

$$\delta G = 0 = \sum \delta n_i g_i + RT \sum \delta n_i \log n_i + RT \sum \delta n_i \\ - RT \sum \delta n_i \log \sum n_i - RT \sum \delta n_i, \quad . . . . (40)$$

or, on account of (33),

$$\sum \nu_i \log n_i - \sum \nu_i \log \sum n_i = - \frac{\sum \nu_i g_i(p, T)}{RT}, \quad . . (41)$$

or since

$$p_i = \frac{n_i}{\sum n_i} p,$$

$$\sum \nu_i \log p_i = - \frac{\sum \nu_i g_i(p, T)}{RT} + \sum \nu_i \log p = \log K_p(p, T).$$

It remains only to calculate the right-hand member. By equations (22) and (23) (p. 536) and equation (36) we have, for  $c_v$  constant,

$$-\sum \nu_i g_i(p, T) + RT \sum \nu_i \log p = -\sum \nu_i u_{i0} - \sum \nu_i c_{vi} T - \sum \nu_i RT \\ + T \log T \sum \nu_i c_{pi} - RT \log p \sum \nu_i + T \sum \nu_i s_{i0} + RT \log p \sum \nu_i \\ = -\sum \nu_i u_{i0} - T \sum \nu_i c_{pi} + T \log T \sum \nu_i c_{pi} + T \sum \nu_i s_{i0}. \quad (43)$$

According to our convention regarding the sign,  $\sum \nu_i u_{i0}$  represents the difference in energy at  $T = 0$  of the system before and after the reaction (written in exothermal form), i.e. the heat of reaction  $q_0^-$  at the absolute zero. After dividing by  $RT$  we note the presence of the combination  $(s_{i0} - c_{pi})/R$ , the "chemical constant"  $i_i$  of the individual gases which we already met in connexion with the vapour pressure curve. Hence, by analogy with equation (30) (p. 538),

$$\sum \nu_i \log p_{i(\text{atm.})} = - \frac{q_0^-(\text{cal})}{4.571 T} + \frac{\log T}{1.985} \sum \nu_i c_{pi} + \sum \nu_i i_i, \quad (44)$$

or using a constant  $B$  which depends only upon the  $i_i$  we have, similar to (16) (p. 535),\*

\* Formulæ (30) and (44) are intended for practical use, while (16) and (44') give a better indication of the theoretical basis.



$$\prod p_i^{\nu_i} = \frac{p_1^{\nu_1} p_2^{\nu_2} \dots}{p_{m+1}^{\nu_{m+1}} p_{m+2}^{\nu_{m+2}} \dots} = Be^{-q^0/RT} T^{\sum \nu_i c_{pi}/R} = K_p(p, T). \quad (44')$$

Equation (44') expresses the celebrated Law of Mass Action, deduced by Guldberg and Waage from kinetic considerations. These investigators, however, were not able to give an expression for the mass action "constant"  $K_p$ , which really depends on  $T$  and  $p$ , and so is a constant only at a given temperature and pressure. The name "Law of Mass Action" comes from the fact that a change in the mass (and hence the partial pressure) of one of the factors in the reaction immediately causes a change in the partial pressures of the others. If, for example, we have an equilibrium between an exothermal compound 1-2, (subscript 3), with its two constituents 1 and 2, and if  $T$  and  $p$  remain constant, the addition of one of the constituents immediately causes the partial pressure of the compound to increase, since the fraction  $p_1 p_2 / p_3$  remains constant. This means that the addition of one of the products of dissociation lowers the degree of dissociation.\*

For many purposes it is more useful to write the Law of Mass action not in terms of partial pressures but in terms of *relative* concentrations  $c_i$  (mol fractions) which are defined by

$$c_i = \frac{n_i}{\sum n_i} \quad . \quad . \quad . \quad . \quad . \quad . \quad (45)$$

Using  $p_i = c_i p$  and  $\nu = \sum \nu_i$ , there follows

$$\prod c_i^{\nu_i} = \frac{c_1^{\nu_1} c_2^{\nu_2} \dots}{c_{m+1}^{\nu_{m+1}} c_{m+2}^{\nu_{m+2}} \dots} = Be^{-q^0/RT} T^{\sum \nu_i c_{pi}/R} p^{-\nu} = K_c(p, T). \quad (46)$$

For reactions which involve no change in the number of mols,  $\nu$  becomes zero and hence  $K_c = K_p$  and the equilibrium does not depend on the pressure. If one compares equation (44') with the vapour pressure formula (16) (p. 535) deduced under the same simplifying assumptions (constancy of specific heats), the similarity becomes evident. This is not fortuitous. In the following sections we shall derive a generalized law of mass action for polyphase systems, in which category the simple process of evaporation of a single substance is also included. As in the case of the vapour pressure formula, we can set up differential relationships for the law of mass action which are free of approximations. Using the mol fractions  $c_i$ , (41) becomes

$$\sum \nu_i \log c_i = - \frac{1}{RT} \sum \nu_i g_i = + \log K_c(p, T). \quad . \quad (47)$$

\* This result, as well as many others, both physical and chemical, comes within the scope of a general law given by Le Châtelier: *If some stress (e.g. a change of pressure, temperature, concentration, &c.) is brought to bear on a system in equilibrium the equilibrium is displaced in the direction which tends to undo the effect of the stress.*

Since  $(\partial g_i / \partial T)_p = -s_i$  (cf. p. 527),

$$\left( \frac{\partial \log K_c}{\partial T} \right)_p = + \frac{1}{RT^2} \sum \nu_i g_i + \frac{1}{RT} \sum \nu_i s_i = \frac{1}{RT^2} \sum \nu_i (u_i + pv_i). \quad (48)$$

But the expression  $\sum \nu_i (u_i + pv_i)$  is the heat of reaction at constant pressure,  $q^-_p$ , by equation (38) (p. 511). Hence

$$\left( \frac{\partial \log K_c}{\partial T} \right)_p = \frac{q^-_p}{RT^2}. \quad \cdot \cdot \cdot \cdot \cdot \quad (48')$$

In the same way, if we denote by  $v$  the change in volume occurring in the reaction (which is determined by holding the pressure constant), i.e.  $\sum \nu_i v_i$ , then we have, since  $(\partial g / \partial p)_T = v$ ,

$$\left( \frac{\partial \log K_c}{\partial p} \right)_T = - \frac{v}{RT}. \quad \cdot \cdot \cdot \cdot \cdot \quad (49)$$

#### 4. Chemical Equilibrium in a System consisting of Dilute Solutions and Ideal Gases.

Those multiple phase systems which contain, besides the gaseous phase, dilute solutions, i.e. phases in which one substance (subscript 0)—the solvent—is present in far larger quantity than any of the others, lend themselves to detailed investigation. We consider a solution to be dilute in the thermodynamical sense if further addition of solvents causes no decrease in volume and if no heat exchange (heat of dilution) is detectable with further dilution. The more nearly these conditions are fulfilled, the better our conclusions will agree with experience. We again examine the equilibrium at a given temperature and pressure and must therefore find the extreme value of the function  $G$ . The separate terms may be calculated as follows: Denoting the solvent by the subscript 0, the internal energy is given by

$$U = U(p, T, n_0, n_1, n_2 \dots). \quad \cdot \cdot \cdot \cdot \cdot \quad (50)$$

On account of the smallness of the mol numbers  $n_1, n_2, \dots$  we can develop this function in a Taylor's series and obtain

$$U = U_0(p, T, n_0) + n_1 \frac{\partial U}{\partial n_1} + n_2 \frac{\partial U}{\partial n_2} + \dots \quad (51)$$

The partial derivatives depend only on the number of mols of the solvent, since the derivatives are taken at  $n_1, n_2, \dots = 0$ . Write  $u_i$  for these derivatives:

$$U = n_0 u_0 + n_1 u_1 + n_2 u_2 + \dots \quad \cdot \cdot \cdot \cdot \cdot \quad (52)$$

In keeping with our original assumptions we neglect the mutual energy of the molecules of the solutes; this corresponds to breaking off the



series after the linear terms. According to Coulomb's Law, the mutual energy of charged particles (e.g. the ions in electrolytes) is large on account of their relatively great charge, even when the particles are far apart. Hence our results will not be applicable to electrolytes.

In the same way we make the analogous abbreviation  $v_i = \partial V / \partial n_i$ , and have

$$V = n_0 v_0 + n_1 v_1 + n_2 v_2 + \dots \quad (53)$$

According to our convention,  $u_i$  and  $v_i$  represent the partial energies and partial volumes respectively of the dissolved substances per mol.

If we were to develop the entropy function in the same way, we would have

$$S = n_0 s_0 + n_1 s_1 + n_2 s_2 + \dots, \quad (54)$$

which would represent the entropy of a system consisting of  $n_0$  mols of solvent *side by side with*  $n_i$  mols of the solutes, but the *mixture* of the substances would not be represented. Hence there must be, in addition, a function of the mol numbers which does not depend on  $T$  or  $p$ . This function is determined if we know its form for any *one* state of the system. According to p. 541, the value of this function is  $-R \sum n_i \log c_i$  for the gaseous state. This value must be retained for all temperatures and pressures. Thus, to a first approximation, we have the same expression as that derived for gases in § 3 (p. 541):

$$G = \sum n_i g_i(T, p) + RT \sum n_i \log c_i. \quad (55)$$

If the solvent does not take part in the reaction we set  $\nu_0 = 0$ , and the chemical equilibrium in a dilute solution is governed by

$$\prod c_i^{\nu_i} = K_c(p, T). \quad (56)$$

The dependence of the equilibrium constant  $K_c$  on temperature is again given by equation (48') (p. 543). There is no variation with pressure, since no change of volume occurs. The most important reaction which takes place in the solution is that of electrolytic dissociation, but in this instance the inter-ionic forces cannot be neglected. This is why strong electrolytes, i.e. those which are highly dissociated, do not obey the law of mass action. In such solutions the interactions give the impression that the dissociation is incomplete, while actually it is practically complete at great dilution. On the other hand, weak electrolytes—especially organic ones—conform very well to the mass action law.

We can now investigate the equilibrium of multiple phase systems, also, provided they consist of ideal gases or dilute solutions. As a special case, of course, individual phases may consist each of a single substance. We must rule out the presence of liquid and solid phases, which comprise several substances in approximately equal concentration, since it is then no longer allowable to terminate the Taylor



expansion with first order terms. We again distinguish the phases by dashes and superscripts, and have for the  $G$  function of the entire system

$$\begin{aligned} G &= \sum n_i' g_i'(p, T) + \sum n_i'' g_i''(p, T) + \dots \\ &\quad + RT(\sum n_i' \log c_i' + \sum n_i'' \log c_i'' + \dots) \\ &= \sum_k \sum_i n_i^{(k)} g_i^{(k)}(p, T) + RT \sum \sum n_i^{(k)} \log c_i^{(k)}. \quad (57) \end{aligned}$$

The determination of the extremal value of  $G$  under the usual auxiliary conditions leads, exactly as above, to the equation of the generalized law of mass action, which also includes transformations between individual phases:

$$\sum_k \sum_i \nu_i^{(k)} \log c_i^{(k)} = -\frac{1}{RT} \sum_k \sum_i \nu_i^{(k)} g_i^{(k)}(p, T) = \log K_c(p, T). \quad (58)$$

The same equations are valid for the derivatives of the function  $\log K_c$  as in the case of the special mass action rule for single-phase systems.

We now consider a two-phase system whose liquid phase is a dilute solution and whose solid phase is the pure solvent. Characterizing the solvent by the subscript 0, we have, for the liquid phase,

$$c_1' = \frac{n_1'}{n_0' + n_1'}, \quad c_0' = \frac{n_0'}{n_0' + n_1'} = 1 - c_1', \quad (59)$$

and for the solid phase

$$c_1'' = 0, \quad c_0'' = 1. \quad (59')$$

The reaction in question is, in the exothermal formulation, the passage of one mol of solvent from the solution into the solid phase, whereby the heat of fusion is liberated as the "heat of reaction"  $q^-$ . Hence the condition of equilibrium is

$$\log(1 - c_1') - \log 1 = \log K_c(p, T). \quad (60)$$

At a given pressure, e.g. atmospheric pressure, this is a relation between the equilibrium temperature  $T$ , i.e. the melting-point, and the concentration of the solution. For the pure solvent alone,  $c_1' = 0$ , and the right member becomes zero, since only  $\log 1$  occurs on the left. If, at constant pressure,  $\log K_c(p, T)$  is developed in a Taylor's series which we terminate with the first term, and if we replace  $\log(1 - c_1')$  by the approximate value  $-c_1'$  we obtain, by equation (48') (p. 543),

$$-c_1' = (T - T_0) \left( \frac{\partial \log K_c}{\partial T} \right)_p = (T - T_0) \frac{q_p^-}{RT_0^2}$$

$$\text{or} \quad T - T_0 = \frac{-n_1'}{n_0' + n_1'} \frac{RT_0^2}{q_p^-} \approx -\frac{n_1}{n_0} \frac{RT_0^2}{q_p^-}. \quad (61)$$

Since the conversion of one mol of liquid solvent to the solid state

liberates heat,  $q_p^-$  is positive, so that *the freezing-point is lowered*. The equation shows the depression of the freezing-point to be proportional to the ratio of the number of mols of dissolved substance to the number of mols of solvent. This is Raoult's Law. The factor of proportionality is determined by the melting-point  $T_0$  of the pure solvent and the heat of fusion  $q_p$ . If several substances are dissolved, we have the sum  $\Sigma n_i/n_0$  in place of  $n_1/n_0$ , as may be verified at once. If, for example, we dissolve a substance which breaks up into two constituents, say a binary  $Z$ - $Z$ -valent electrolyte, the depression of the freezing-point must be twice as great as for the same number of mols of a substance which does not decompose. In reality the double value is attained only when the dilution is very great, from which we already concluded that equilibrium between molecules and ions exists, especially since the degree of dissociation as determined by the lowering of the freezing-point agrees well with the values obtained from electrolytic conduction. Actually the dissociation of strong electrolytes is much greater than that to be expected on this ground; the deviations from the value corresponding to complete dissociation are for the most part attributable to the electrical forces operating between the particles (cf. the following section).

For substances which do not break up (non-ionogens) we obtain the molecular weight from the lowering of the freezing-point. It is readily calculated from (61) that one molecular weight of such a substance dissolved in one litre of water causes the freezing-point to drop to  $-1.86^\circ \text{C}$ .

The same calculations hold also for the *elevation of the boiling-point*, the pressure remaining constant. The sign of  $T - T_0$  is reversed, since the solution is now the phase of lesser energy. The calculation shows that one mol of a non-ionogen dissolved in one litre of water raises the boiling-point to  $100.52^\circ \text{C}$ .

Finally, we wish to derive the formula for the osmotic pressure of a solution. Consider a system composed of two liquid phases, the solution and the solvent. They are prevented from mixing by a separating wall. Let this wall be selectively permeable to the solvent. The wall may be considered non-existent for a substance whose free passage it permits; hence there must be the same number of mols of this substance per cubic centimetre on both sides of the membrane, as in the case of a gas. But in the system under consideration this is certainly not the case; for, besides the solvent, the dissolved substance is also present on the solution side, so that the solvent will continue to pass through the membrane until there is no longer any difference in concentration on the two sides. Thus with a solution of finite concentration and equal pressure on the two sides, equilibrium cannot exist. However, by placing the solution side under higher pressure, the passage of solvent can be prevented.



The two phases are thus under different pressures. A system of this kind is not included in our previous considerations, and the conditions for equilibrium must be derived anew. We wish to examine the equilibrium at a given temperature, a given pressure and a given pressure difference, which we call the osmotic pressure. Hence a virtual displacement of equilibrium is free to cause only a change in the number of mols in each phase, and it is evident that the condition of equilibrium  $\delta G = 0$  again leads to the equation

$$\Sigma \Sigma \nu_i^k \log c_i^k = -\frac{1}{RT} \Sigma \Sigma \nu_i^k g_i^k.$$

The difference is that the functions  $g_i^k$  refer to different pressures. In the case at hand, the reaction consists of the transfer of one mol of solvent to the solution side. To distinguish the solution from the solvent, the former is denoted by a prime. Then

$$\begin{aligned} 1 \log c_0' - 1 \log 1 &= \log(1 - c_1') \approx -c_1' \\ &= \frac{1}{RT} g_0(p, T) - \frac{1}{RT} g_0'(p + p_{\text{osm}}, T). \quad (62) \end{aligned}$$

Developing the right member in a Taylor's series which we terminate with the first term, and using the relation (31) (p. 527), we obtain

$$\begin{aligned} c_2' &\approx \frac{n_1'}{n_0'} = \frac{1}{RT} \left( \frac{\partial g}{\partial p} \right)_T p_{\text{osm}} = \frac{p_{\text{osm}} v}{RT} \\ \text{or} \quad p_{\text{osm}} &= \frac{n_1' RT}{n_0' v} \approx \frac{n_1' RT}{V'}. \quad (63) \end{aligned}$$

Thus the osmotic pressure is equal to the pressure which the same number of mols of the substance would exert if they were in the form of a gas occupying the volume of the solution at the same temperature. This was first shown by van't Hoff. The law will be deduced from another point of view on p. 585.

## 5. Thermodynamic Equilibrium of Dilute Solutions of Strong Electrolytes.

As already mentioned, the older theory recognized the existence of an equilibrium between the molecules and the ions of strong electrolytes. However, for strong electrolytes, i.e. those which are good conductors, the equilibrium does not satisfy the Law of Mass Action. The reason is that it is not permissible to neglect the inter-ionic electrical forces in formulating the thermodynamic functions. The previously employed assumption that the individual particles may be treated like the molecules of an ideal gas is evidently not justified here. We must add to the previously used internal energy  $U_{\text{id}}$  an



electrical contribution which, by the method of Debye, may be computed without undue difficulty.

In the case of electrical conduction the situation is more complicated, and so we refrained from going through the very cumbersome calculations when we treated this phenomenon on p. 423.

The entropy of the system is also altered; as a result the function  $G$ , which determines all equilibria at a given temperature and pressure, also has an added term:

$$G = G_{id} + G_{el} = U + pV - TS. \quad . \quad . \quad . \quad (64)$$

For condensed systems the term  $pV$  is not of consequence, and it is not necessary to distinguish between  $(\partial G/\partial T)_p$  and  $(\partial G/\partial T)_v$ ; hence by equation (33) (p. 528) we have the following differential equation for  $G$ :

$$G = U + T \frac{dG}{dT}.$$

Since this equation is satisfied by  $G_{id}$ , the added electrical term  $G_{el}$  must also satisfy it; this serves to determine  $G_{el}$  from  $U_{el}$ , for it follows by integration of

$$G_{el} = U_{el} + T \frac{dG_{el}}{dT} \quad . \quad . \quad . \quad . \quad . \quad (65)$$

that

$$G_{el} = -T \int \frac{U_{el}}{T^2} dT.* \quad . \quad . \quad . \quad . \quad (66)$$

This shows that most of the work required to find the function  $G$  is done if the electrical energy  $U_{el}$  is known as a function of the temperature. This will now be determined for a binary  $Z$ - $Z$ -valent electrolyte. The extension of the calculation to more complicated cases is not difficult. Let us consider a single positive ion. As mentioned on p. 423, there will be—averaged in time—more negative ions in its neighbourhood than positive. The average distribution of charge may then be replaced by a continuous space charge determined as follows: According to Boltzmann's Principle (cf. p. 588), the number of particles possessing an energy  $u$  is proportional to

$$e^{-u/kT},$$

\* The integration is effected by starting with the homogeneous equation

$$T \frac{dG_{el}}{dT} - G_{el} = 0,$$

whose integral is  $G_{el} = aT$ . We now vary the parameter, putting  $G_{el} = T \cdot v(T)$  and obtaining the differential equation

$$T^2 \frac{dv}{dT} + U_{el} = 0$$

for  $v$ . The integral is

$$v = - \int \frac{U_{el}}{T^2} dT.$$

where  $k$  is the gas constant per molecule,  $k = R/L$ . At a place where the potential is  $\psi$  the charge  $Ze$  has the energy  $Ze\psi$ , so that the number of positive ions in a volume element  $dV$  at that point is

$$dN_+ = Ce^{-Ze\psi/kT} dV. \quad . \quad . \quad . \quad . \quad (67)$$

The constant  $C$  is determined from the fact that uniform distribution must result for  $\psi = 0$ ;  $C$  is thus the average number of positive ions per cubic centimetre, which we denote by  $N_+$ . In the same way, the number of negative ions in the same volume element is

$$dN_- = N_- e^{+Ze\psi/kT} dV. \quad . \quad . \quad . \quad . \quad (68)$$

For a  $Z$ - $Z$ -valent electrolyte  $N_+ = N_- = N$ . Hence the space charge density at a point where the potential is  $\psi$  is

$$\rho = -2NZe \sinh \frac{Ze\psi}{kT}. \quad . \quad . \quad . \quad . \quad (69)$$

Now  $\rho$  and  $\psi$  are connected by Poisson's equation (p. 278 *et seq.*):

$$\Delta\psi = -\frac{4\pi\rho}{D} = \frac{8\pi NZe}{D} \sinh \frac{Ze\psi}{kT}, \quad . \quad . \quad . \quad (70)$$

where  $D$  is the dielectric constant of the solvent. The value of the potential at a distance of one ionic radius from the point in question is such that we may develop the term  $\sinh(Ze\psi/kT)$  and omit the higher powers. In this way we obtain

$$\Delta\psi = \frac{8\pi NZ^2 e^2}{DkT} \psi = \kappa^2 \psi, \quad . \quad . \quad . \quad . \quad (71)$$

where

$$\kappa = \sqrt{\frac{8\pi NZ^2 e^2}{DkT}}.$$

The quantity  $\kappa$ , which has the dimensions of the reciprocal of a length, appears in all the subsequent computations. Its meaning becomes clear if we digress for a moment from the problem in hand, viz. that of the distribution of charge round a central point, and consider the simpler linear case of a plate at the potential  $\psi_0$  dipping into a liquid electrolyte. The differential equation is then

$$\frac{d^2\psi}{dx^2} = \kappa^2 \psi, \quad . \quad . \quad . \quad . \quad (72)$$

and its integral is \*

$$\psi = \psi_0 e^{-\kappa x}. \quad . \quad . \quad . \quad . \quad (73)$$

Thus  $1/\kappa$  is the distance in which the potential—and hence, by Poisson's equation, the negative space charge—has diminished to  $1/e$  times its original value, i.e.  $1/\kappa$  is a measure of the thickness of the compensating layer. The smaller  $N$  is, i.e. the more dilute the solution, the larger is the quantity  $1/\kappa$ .

\* The letter  $e$  is used in two ways here, but there is no possibility of confusion, as the base of natural logarithms almost always carries an algebraic exponent.

For the spherically symmetric case which interests us here the differential equation becomes

$$\frac{d^2\psi}{dr^2} + \frac{2}{r} \frac{d\psi}{dr} = \kappa^2\psi. \quad . \quad . \quad . \quad . \quad . \quad (74)$$

As may be verified by substitution, the general solution is

$$\psi = A \frac{e^{-\kappa r}}{r} + B \frac{e^{+\kappa r}}{r}. \quad . \quad . \quad . \quad . \quad . \quad (75)$$

Since the potential must vanish at infinity  $B = 0$ . The constant  $A$  is determined by the fact that the potential in the immediate neighbourhood of the selected ion must be that of the ion alone. The resulting value of  $A$  is then  $Ze/D$ , and the solution becomes

$$\psi = \frac{Ze}{D} \cdot \frac{e^{-\kappa r}}{r} = \frac{Ze}{Dr} - \frac{Ze}{Dr} (1 - e^{-\kappa r}). \quad . \quad . \quad . \quad (76)$$

The second form shows how the potential is made up of one part attributable to the central ion and another part due to the presence of the surrounding ions of opposite sign—the “ion cloud”. At  $r = 0$ , i.e. at the centre of the selected ion, the cloud generates the potential

$$\psi_0 = - \lim_{r \rightarrow 0} \frac{Ze}{Dr} (1 - e^{-\kappa r}) = - \frac{Ze\kappa}{D}. \quad . \quad . \quad (77)$$

This is equal to the work required to bring a unit charge from infinity to the middle of the spherical cloud. Hence the potential energy possessed by the central positive ion is

$$u_+ = + Ze\psi_0 = - \frac{Z^2e^2\kappa}{D}. \quad . \quad . \quad . \quad . \quad (78)$$

Similarly, for a negative ion we would have

$$u_- = - \frac{Z^2e^2\kappa}{D}.$$

Since each ion is counted twice over—once as the central ion, once as part of the ion cloud—the total energy is

$$U_{el} = - \frac{1}{2} \frac{2NVZ^2e^2\kappa}{D} = - \frac{NVZ^2e^2}{D} \sqrt{\frac{8\pi NZ^2e^2}{DkT}}. \quad . \quad (79)$$

Substituting this in equation (66) (p. 548) and performing the integration,

$$G_{el} = - \frac{2}{3} \frac{NVZ^2e^2\kappa}{D}, \quad . \quad . \quad . \quad . \quad (80)$$



or, for  $n_1$  mols of solute,

$$G_{el} = -\frac{2}{3} \frac{n_1 LZ^2 e^2 \kappa}{D}, \quad . . . . . (80')$$

where  $L$  is Loschmidt's (Avogadro's) Number. Then the Gibbs Potential of the electrolytic solution at complete dissociation is

$$G = n_0 g_0 + 2n_1 g_1 + RT n_0 \log c_0 + 2RT n_1 \log c_1 - \frac{2}{3} n_1 \frac{LZ^2 e^2 \kappa}{D}, \quad (81)$$

$$\text{where } \kappa = \sqrt{\frac{8\pi N Z^2 e^2}{D k T}} = \sqrt{\frac{8\pi n_1 LZ^2 e^2}{V D k T}} \approx \sqrt{\frac{8\pi n_1 LZ^2 e^2}{n_0 v_0 D k T}}. \quad (82)$$

The last form shows that  $\kappa$  depends on the number of mols of the solvent. Once the characteristic function  $G$  is determined, all else is merely a matter of computation.

We now calculate the difference between the lowering of the freezing-point in this case and that for complete dissociation without electrical interaction. The process consists of the conversion of one mol of solvent from the liquid state to the solid state. The only quantities which vary are the numbers of mols of the solvent in each phase; it is to be remembered, however, that these enter into the quantity  $\kappa$ . We shall not yet replace the variations  $\delta n_i$  in the equilibrium formula by the conversion numbers  $\nu_i$  which are proportional to them. Indicate the liquid phase by one dash attached to the letters and the solid phase by two dashes. Then we have, exactly as in § 4 (p. 545),

$$\begin{aligned} & \delta n_0'' \log c_0'' + \delta n_0' \log c_0' \\ &= -\frac{1}{RT} \left\{ \delta n_0'' g_0'' + \delta n_0' g_0' - \frac{2}{3} \frac{n_1 LZ^2 e^2}{D} \frac{\partial \kappa}{\partial n_0'} \delta n_0' \right\}. \quad (83) \end{aligned}$$

Noticing that

$$\delta n_0'' = -\delta n_0'; \quad c_0'' = 1; \quad c_0' = 1 - 2c_1'; \quad \log c_0' \approx -2c_1', \quad (84)$$

we obtain

$$-2c_1' = -\frac{1}{RT} \left\{ g_0' - g_0'' - \frac{2}{3} \frac{n_1 LZ^2 e^2}{D} \frac{\partial \kappa}{\partial n_0'} \right\}. \quad (85)$$

If, as in § 4, we develop the functions  $g_0'$  and  $g_0''$  at the point  $T_0$ , i.e. the freezing-point of the pure solvent, and if we remember that

$$\left( \frac{\partial g_0'}{\partial T} \right)_p = -s_0'; \quad \left( \frac{\partial g_0''}{\partial T} \right)_p = -s_0''; \quad s_0' - s_0'' = \frac{q_p^-}{T_0}, \quad (86)$$

we obtain

$$2c_1' = -\frac{\Delta T q_p^-}{RT_0^2} - \frac{2}{3} \frac{n_1 LZ^2 e^2}{D RT_0} \frac{\partial \kappa}{\partial n_0'} \quad . . . (87)$$

If the electrical forces were not taken into account we would have for the lowering of the freezing-point, at complete dissociation,

$$\Delta T_{\text{id}} = - \frac{2c_1' RT_0^2}{q_p^-}, \quad \text{i.e.} \quad 2c_1' = - \frac{\Delta T_{\text{id}} q_p^-}{RT_0^2}. \quad (88)$$

Using this expression to compute the relative departure from the present value we obtain

$$\frac{\Delta T_{\text{id}} - \Delta T}{\Delta T_{\text{id}}} = 1 - \frac{\Delta T}{\Delta T_{\text{id}}} = - \frac{1}{3} \frac{n_1' LZ^2 e^2}{DRT_0 c_1'} \frac{\partial \kappa}{\partial n_0'}. \quad (89)$$

It remains only to compute the value of  $\partial \kappa / \partial n_0'$ . From the approximate expression

$$\kappa = \sqrt{\frac{8\pi n_1' LZ^2 e^2}{n_0' v_0 D k T_0}}$$

we have at once

$$\frac{\partial \kappa}{\partial n_0'} = -\frac{1}{2} \frac{\kappa}{n_0'}, \quad \dots \dots \dots (90)$$

and since  $n_1'/n_0' = c_1'$  and  $R = kL$ ,

$$1 - \frac{\Delta T}{\Delta T_{\text{id}}} = \frac{1}{6} \frac{LZ^2 e^2 \kappa}{RT_0 D} = \frac{\sqrt{2\pi}}{3} \sqrt{\frac{n_1'}{V_0'}} \frac{L^2 Z^3 e^3}{D^{3/2} (RT_0)^{3/2}}. \quad (91)$$

We then have the following result: The relative departures from the value of the depression of the freezing-point for complete dissociation and negligible electrical interaction are directly proportional to the square root of the molar concentration  $n_1'/V_0'$ , inversely proportional to the  $3/2$  power of the dielectric constant and—for  $Z$ - $Z$ -valent electrolytes—directly proportional to the cube of the valence. This relationship is well substantiated experimentally for very dilute solutions.

A refinement of the calculation must take into account the finite ionic radius; this has been done in the complete theory of Debye and Hückel.

## CHAPTER XXXIII

### THE NERNST HEAT THEOREM

#### 1. Free Energy as a Measure of Chemical Affinity; Determination of this Quantity for the Galvanic Cell.

In the "classical" thermodynamics, which employs only the two fundamental laws discussed in previous chapters, the absolute value of the entropy remains undetermined; as a result, the values of the functions  $F$  and  $G$  are fixed only as far as a linear function of  $T$ . This indeterminacy becomes very inconvenient in certain applications. Thus we saw in the preceding chapter (p. 534) that it is not possible to calculate the absolute value of the logarithm of the vapour pressure from a knowledge of the specific heat and heat of vaporization—that we required, in addition, a knowledge of the "*chemical constant*", which is closely connected with the entropy constant. But the latter is determined by the two laws of thermodynamics only to the extent that a value once given must be preserved in all changes of the system. On the other hand, the Nernst theorem leads to a definite statement regarding the entropy constant of solid and liquid substances. The *calculation* of the chemical constant of a gas can be performed only with the help of the Quantum Theory. It is through a knowledge of the entropy constant, or of the chemical constant, that the equilibrium formulæ of the preceding chapter become really of practical use. Only when this is known can a chemical equilibrium be computed theoretically.

Nernst arrived at his heat theorem through the problem of calculating chemical affinity. It was a considerable time before a practical measure of the somewhat anthropomorphic concept of the chemical affinity of two substances—or, as we say to-day, of a chemical reaction—was found. One of the foremost demands to be made of such a quantity would be that it tell in which direction the reaction proceeds. Berthelot took the heat of reaction as a measure of affinity, and—as a matter of fact—at room temperature most reactions are accompanied by the evolution of heat. Nevertheless, there are some very common processes—like the dissolving of most salts in water—which are accompanied by cooling, and which are therefore endothermal. It was van't Hoff who first recognized that for reactions at constant volume the free energy  $F$  is the proper measure of affinity,



while for reactions proceeding at constant pressure it is the Gibbs Potential  $G$ . These quantities are a minimum at equilibrium; the greater the values of these functions compared with the minimum values, the farther the system is from equilibrium—hence the more readily will the reaction occur. The free energy of a chemical reaction proceeding at constant pressure is easily determined experimentally by using the reaction in a simple reversible galvanic cell and measuring the electromotive force of the cell. For condensed systems, i.e. those not containing the gaseous phase, it is not necessary to distinguish between isometric and isobaric reactions and hence between  $F$  and  $G$ . Suppose the chemical change in this cell takes place by the migration of an ion bearing a charge of  $Z$  units. Then when one mol has been transported, a charge  $Z\mathfrak{F}$  has passed. The Faraday equivalent charge is denoted by  $\mathfrak{F}$  here to avoid confusing it with the free energy. For a difference of potential  $\Phi$ , the work done is given by  $Z\mathfrak{F}\Phi$ , and this is equal to the change in free energy:

$$Z\mathfrak{F}\Phi = f_1 - f_2. \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

Now  $f$  satisfies the differential equation

$$f = u + T \left( \frac{\partial f}{\partial T} \right)_v,$$

so that

$$f_1 - f_2 = u_1 - u_2 + T \left( \frac{\partial f_1}{\partial T} \right)_v - T \left( \frac{\partial f_2}{\partial T} \right)_v, \quad . \quad . \quad (2)$$

and hence by (1),

$$\Phi = \frac{q_v^-}{Z\mathfrak{F}} + T \left( \frac{\partial \Phi}{\partial T} \right)_v. \quad . \quad . \quad . \quad . \quad . \quad (3)$$

This important relationship between the potential difference, the heat of reaction of the chemical process and the temperature coefficient of the potential difference, may be derived also from the equation  $TdS = dU + dW$  if the external work is expressed in terms of the electrical quantities (cf. *Ex.* 125, p. 528). Since it is usual to express  $\Phi$  in volts and  $\mathfrak{F}$  in coulombs, the heat of reaction must be expressed in joules.

By examination of the empirical data, it is found that the electromotive force of most cells at room temperature is fairly accurately given by the simple formula

$$\Phi = \frac{q_v^-}{Z\mathfrak{F}} = \frac{u_1 - u_2}{Z\mathfrak{F}}$$

obtained by setting the change in free energy equal to that of the total energy. The lower the temperature at which the measurements are taken, the better the agreement is found to be. One would con-

clude from this that at absolute zero the differences  $f_1 - f_2$  and  $u_1 - u_2$  are equal, i.e. that Berthelot's Principle is strictly true at that temperature. By equation (2), this means that

$$\frac{\partial f_2}{\partial T} - \frac{\partial f_1}{\partial T}$$

remains finite for  $T = 0$ , or else that it becomes infinite to so low an order that the limit

$$T \left( \frac{\partial f_2}{\partial T} - \frac{\partial f_1}{\partial T} \right)$$

vanishes. But since the agreement is still good at relatively high temperatures, Nernst made the hypothesis that the curves  $f_2(T) - f_1(T)$  and  $u_2(T) - u_1(T)$  not only intersect at  $T = 0$  but touch each other there. Corresponding to the way in which he arrived at his theorem, Nernst made a statement concerning only the difference of the functions  $f$  and  $u$  before and after the reaction, but Planck extended the hypothesis by assuming that it holds for  $f_1$  and  $f_2$  separately, i.e. that

$$\left. \begin{aligned} \lim_{T=0} f(T) &= \lim_{T=0} u(T) \\ \lim_{T=0} \left( \frac{\partial f}{\partial T} \right)_v &= \lim_{T=0} \left( \frac{\partial u}{\partial T} \right)_v \end{aligned} \right\} \cdot \cdot \cdot \cdot (4)$$

As we shall see in the following section, this serves to determine the entropy constant of pure, condensed substances.

## 2. Formulation of the Nernst Heat Theorem.

We obtain a new form of the Nernst theorem by solving the differential equation of free energy (deduced on p. 527) for  $(\partial f / \partial T)_v$ :

$$\left( \frac{\partial f}{\partial T} \right)_v = \frac{f - u}{T} \cdot \cdot \cdot \cdot (5)$$

Since  $f$  becomes zero when  $T = 0$ , the expression  $\partial f / \partial T$  assumes the indeterminate form  $0/0$  under these conditions. The usual means of evaluating the fraction is to differentiate numerator and denominator separately and then substitute  $T = 0$ . This yields

$$\left( \frac{\partial f}{\partial T} \right)_0 = \left( \frac{\partial f}{\partial T} \right)_0 - \left( \frac{\partial u}{\partial T} \right)_0 \cdot \cdot \cdot \cdot (6)$$

Now, by the Nernst-Planck relation, the entire right member is zero, so

that each term on the right must be separately zero in this case. The character of the function is represented in fig. 1.

According to equation (30), p. 526,

$$\left(\frac{\partial f}{\partial T}\right)_v = -S, \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

from which it follows that the entropy  $S$  also vanishes at the absolute zero.

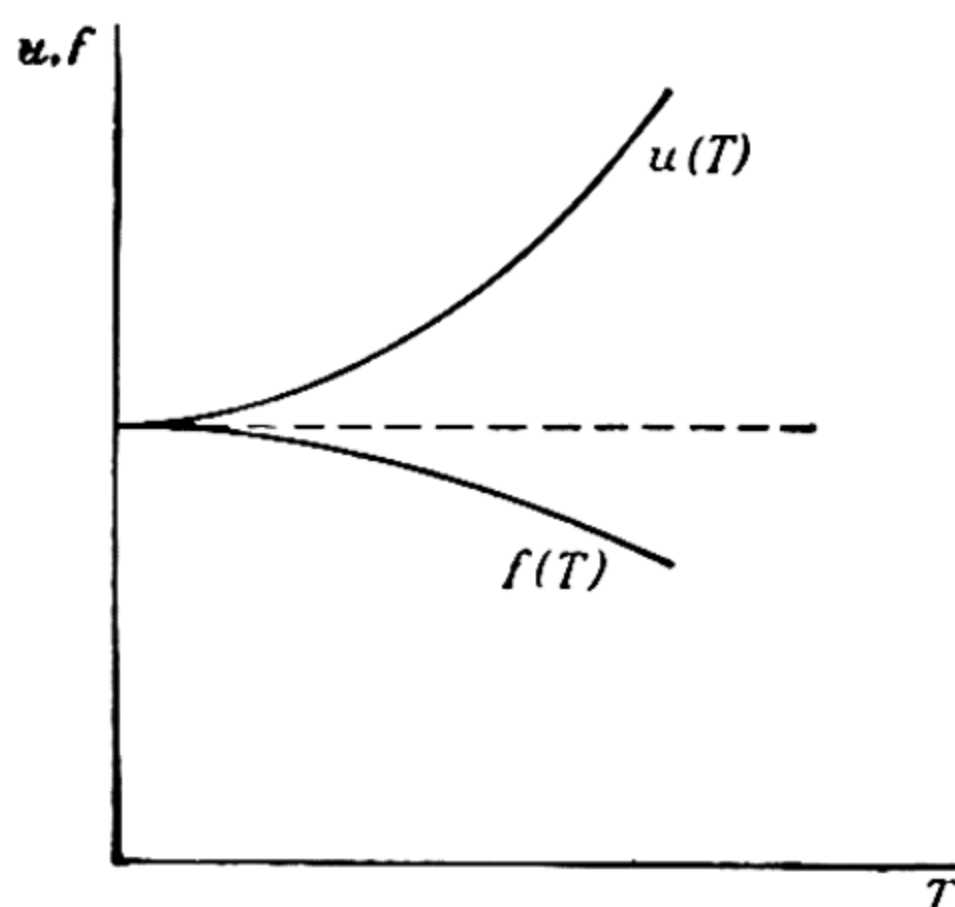


Fig. 1

We may therefore state the Nernst theorem in this way: *At the absolute zero of temperature the entropy of a pure solid or liquid has the value zero.* This statement applies only to pure substances, since for mixtures there must be an additional term covering the increase of entropy which accompanies the process of mixing. We must also except gases, as long as we retain the classical gas equation, for this equation yields an infinite entropy constant at  $T = 0$  if the value of  $S$  is to remain finite. In reality, it follows from the new statistics (p. 626 *et seq.*) that the equation of state of a gas is subject to a correction which preserves the general validity of the Nernst theorem. In particular, there can be no difference in the entropy of two different modifications of a substance at the absolute zero. For instance, a paramagnetic material must have the same zero-point entropy whether it is magnetized or not; i.e. it must then behave like a permanent magnet even without the field.

In a statistical interpretation, the Nernst theorem affirms that a body must be in its greatest state of order at the absolute zero. It is precisely here that the theorem has a limitation as far as its practical meaning is concerned. To be sure, the state of maximum order is that of equilibrium at  $T = 0$ , but as to the amount of time required for the attainment of that equilibrium, thermodynamics has nothing to say. Since the motions of ions and of molecules dwindle at low temperatures, states of non-equilibrium can persist indefinitely. For example, glasses



have a state of higher order—the crystalline—and yet a piece of glass is still glass at  $T = 0$ ! Circumstances are otherwise in connexion with the motion of electrons and the orientation of the magnetic moments connected with them. Their motion is not hindered by any increase of internal friction with falling temperature. For this reason the consequences for magnetization are of especial importance.

### 3. Consequences for the Specific Heats and Temperature Coefficients The Chemical Constant of a Gas.

It follows from

$$ds = \left( \frac{\partial s}{\partial T} \right)_v dT + \left( \frac{\partial s}{\partial v} \right)_T dv = \frac{1}{T} (du + p dv) \quad \cdot \quad (8)$$

that at constant volume

$$\left( \frac{\partial s}{\partial T} \right)_v = \frac{1}{T} \left( \frac{\partial u}{\partial T} \right)_v = \frac{c_v}{T}, \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad (9)$$

and

$$\frac{\partial^2 s}{\partial T \partial v} = \frac{1}{T} \frac{\partial c_v}{\partial v}. \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad (10)$$

This is a relation we shall need later. Integrating (9) from 0 to  $T$  we obtain, by Nernst's theorem,

$$s = \int_0^T \frac{c_v}{T} dT. \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad (11)$$

In order that the integral may converge, the specific heat (for condensed systems it is unnecessary to distinguish between  $c_v$  and  $c_p$ ) must approach zero sufficiently rapidly as the temperature diminishes. Actually, the theory of the specific heats of solids (p. 612) gives a decrease as the third power of  $T$ . On the other hand, the value given by the classical statistics for the specific heat of a gas is different from zero even at  $T = 0$ . Hence we must leave gases out of consideration as long as we take the classical statistics to be correct (see preceding section). It follows further from the Nernst Heat Theorem that all temperature coefficients of condensed substances vanish at the absolute zero. We shall demonstrate this for the isometric pressure coefficient. It follows from

$$s = - \left( \frac{\partial f}{\partial T} \right)_v \quad \text{and} \quad p = - \left( \frac{\partial f}{\partial v} \right)_T \quad \cdot \quad \cdot \quad \cdot \quad (12)$$

that  $(\partial p / \partial T)_v$  is given by

$$\left( \frac{\partial p}{\partial T} \right)_v = \left( \frac{\partial s}{\partial v} \right)_T = \int_0^T \frac{1}{T} \frac{\partial c_v}{\partial v} dT, \quad \cdot \quad \cdot \quad \cdot \quad (13)$$

and from equations (10) and (13),

$$\left(\frac{\partial p}{\partial T}\right)_v = \int_0^T \frac{\partial^2 s}{\partial T \partial v} dT = \int_0^T \left(\frac{\partial^2 p}{\partial T^2}\right)_v dT = \left(\frac{\partial p}{\partial T}\right)_v - \left(\frac{\partial p}{\partial T}\right)_{v, T=0},$$

whence

$$\left(\frac{\partial p}{\partial T}\right)_{v, T=0} = 0. \quad \cdot \cdot \cdot \cdot \cdot \cdot (14)$$

As regards gases, it has already been mentioned that theories have been developed in recent years which yield the usual gas equation as an approximation at higher temperatures, while in the neighbourhood of absolute zero neither the equation of state nor the constancy of the specific heat of a monatomic gas is retained. This is called the degeneration of a gas. Since all gases condense at temperatures considerably above the absolute zero, attempts to verify this degeneracy experimentally have not yet succeeded. We shall not take up these theories (cf. p. 628) at this point, but shall pursue another line of inquiry; this will not yield the value of the entropy at absolute zero, but will determine the entropy constant at finite temperatures. For this purpose we consider the equilibrium between a gas and its condensed liquid, for which system the vapour pressure formula (29) (p. 538) is valid. Here the entropy constant  $s_{2,0}$  of the liquid is to be set equal to zero. Again, we imagine the measurements performed at very low temperatures, where the integrals contribute nothing. This is because the specific heats of the liquid vanish as  $T^3$  and the part of the specific heat of the gas which depends on  $T$  also approaches zero rapidly, and all that is left for any gas is the specific heat of a monatomic gas as the part  $c_{pc}$  which is independent of the temperature. The vapour pressure equation is then

$$\log p = -\frac{q_0}{RT} + \frac{c_{pc}}{R} \log T + \frac{s_{10} - c_{pc}}{R}. \quad \cdot \cdot (15)$$

Thus, by measuring the vapour pressure, the chemical constant

$$i_{\text{abs}} = \frac{s_{10} - c_{pc}}{R} \quad \cdot \cdot \cdot \cdot \cdot (16)$$

(and therefore the entropy constant) of the gas may be determined.\* On the other hand, the theoretical calculation of this quantity is possible only with the aid of quantum statistics.

\* The subscript "abs" is appended to  $i$  to denote that this is not the usual  $i$  in which the logarithms of the conversion factors are also included (cf. p. 538).

## 4. Unattainability of the Absolute Zero.

It is a consequence of the Nerst heat theorem that the absolute zero is an asymptotic point—it may be approached indefinitely but never strictly reached. Since many interesting phenomena are encountered in the small range between, say,  $1^\circ$  absolute and 0, it would be better to use a logarithmic scale of temperature in this region, for the scale itself would then have an unattainable zero.

In order to demonstrate the unattainability, we consider the most feasible means of cooling—that of adiabatic demagnetization discussed earlier. Assume that the entropy-temperature curves have the form shown in fig. 2*a*, which is contrary to the Nernst theorem. In any case,

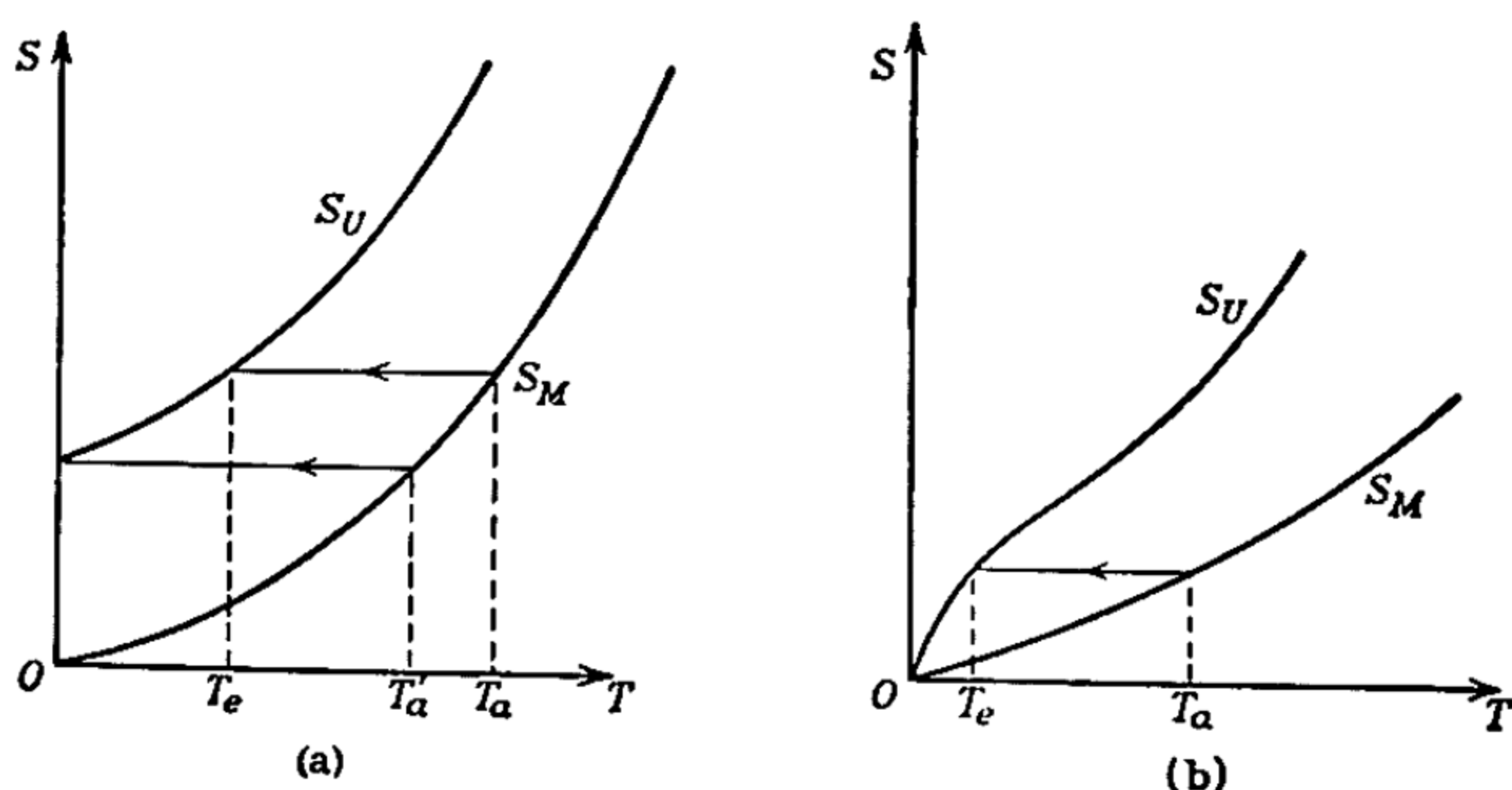


Fig. 2

at higher temperatures the entropy curve  $S_U$  of the unmagnetized\* body must lie above the curve  $S_M$  of the magnetized body. The reason is that entropy, as will be explained in the following chapter, is a measure of the disorder; and this is certainly greater in the unmagnetized state, where the elementary magnets are oriented at random. Now assume demagnetization to take place adiabatically and reversibly at a temperature  $T_a$ . The process will be isentropic, for if  $\delta Q$  is zero in the reversible process, then  $\delta S$  must also be zero. In the diagram, then, we proceed along a horizontal line. The point of intersection with  $S_U$  determines the final temperature. If we were to start from a temperature  $T'_a$  or lower, we would attain the absolute zero by the time the  $S_U$  curve is reached. However, according to the Nernst theorem there can be no entropy difference between the two forms at the zero point,

\* By "magnetized" we shall mean, more accurately, a paramagnetic body in a magnetic field. The coincidence of  $S_U$  and  $S_M$  at  $T = 0$  means that a body at zero behaves as though a field were present, i.e. it becomes ferromagnetic.



and so the curves must run as shown in fig. 2*b*, in which case the zero point cannot be reached. Starting from a temperature  $T$ , the steepness of descent of the curve  $S_v$  determines how close to  $T = 0$  one may come.

No matter what adiabatic process is considered, the same conclusion results. If, for instance, in analogy with the expansion of a gas, we seek to cool a solid \* by sudden extension, we get similar curves for two different pressures, since at any given pressure  $s$  must be zero for  $T = 0$ .

Despite the fact that the unattainability of the absolute zero can thus be deduced from the Nernst law, the two must not be considered equivalent formulations. This is the case only if we add the experimental fact that the specific heats of all substances approach zero at least as  $T^3$  (see p. 612).

\* At low temperatures gases no longer exist as such.

## **PART VI**

### **THE THEORY OF HEAT. STATISTICAL PART**





## CHAPTER XXXIV

### THE ELEMENTARY KINETIC THEORY OF MATTER

#### 1. Bernoulli's Formula; Boyle's Law.

As already indicated on p. 496, an ideal gas is defined from the atomistic point of view as an aggregate of particles which have no extension, and between which no forces operate. These particles are in motion amongst each other with all possible velocities. The apparently steady pressure exerted by the gas is due to the myriad impacts of these particles on the walls, from which they rebound elastically. Collisions between molecules are not possible for an ideal gas, since the particles have no extension. We wish to calculate the pressure. Resolve the velocity  $c$  of a molecule into its rectangular components  $c_x, c_y, c_z$ . Consider a portion of a wall normal to the  $x$ -axis. Each time a molecule collides with this wall it suffers a change in the  $x$ -component of its momentum amounting to  $-2mc_x$ , while the other components are not altered by the elastic collision. But according to Newton's Third Law, the impulse communicated to the wall is also of this magnitude. The impacts follow one another in such rapid succession that the series of impulses is felt as a constant force. If each impact lasts a short time  $\Delta t$ , the average force  $\overline{F}_x'$  which the wall applies to the molecule is given by (p. 108):

$$\overline{F}_x' = \frac{1}{\Delta t} \int_0^{\Delta t} F_x' dt = - \frac{1}{\Delta t} 2mc_x. \quad \dots \quad (1)$$

According to Newton's Third Law, the force exerted by the molecule on the wall is

$$\overline{F}_x = - \overline{F}_x' = + \frac{1}{\Delta t} 2mc_x. \quad \dots \quad (1')$$

If, instead of a single molecule,  $\Delta \nu$  of them strike the wall in the time  $\Delta t$ , the force is multiplied by  $\Delta \nu$ . The pressure is given by the force on unit area. If the number of molecules striking each square centimetre in the time  $\Delta t$  is  $\Delta N$ , the pressure is

$$p = \frac{\Delta N}{\Delta t} 2mc_x. \quad \dots \quad (2)$$

The quotient  $\Delta N/\Delta t$  is the number of molecules striking each unit area of the wall per second.

All possible velocities occur, with respect to both direction and magnitude. With regard to the former, there is no uniquely indicated direction in space; we may therefore consider that, on the average, one-third of all the molecules move in the direction of each axis. Of each of these groups, half move in the direction of the positive axis, half in the negative direction. Further, we assume here that all the molecules have the same speed. In the next second, all those molecules within a cylinder of base unity and altitude  $c$  normal to the wall which are moving *toward* the wall will collide with it, for a molecule distant  $c$  from the wall and approaching it will just reach it in one second, while all those farther away will not. If there are  $N$  molecules in each cubic centimetre,  $N/6$  of these move toward the wall which is perpendicular to the  $x$ -axis. In all,  $Nc/6$  molecules collide with this wall each second, so that

$$p = \frac{Nmc^2}{3}. \quad . . . . . (3)$$

Since all possible velocities occur,  $c^2$  is to be replaced by its average value  $\overline{c^2}$ , the mean square velocity. The square root of this quantity—the so-called “root mean square” velocity—is only approximately equal to the mean velocity (cf. p. 593).

Let us add here a more exact deduction of formula (3)—one that takes into account the distribution of the velocities of the molecules. In reality, there is no molecule that moves with exactly a given speed in a given direction at any instant. Rather, we must assume that in each cubic centimetre there are  $d^3N$  molecules \* whose velocities have magnitudes falling between  $c$  and  $c + dc$ , and directions lying within a solid angle  $d\Omega$ .

Since all directions are on a par with each other, the number of molecules whose directions fall within a solid angle  $d\Omega$  will be proportional to the size of this angle; the number of molecules whose speeds fall between  $c$  and  $c + dc$  will, however, depend on  $c$  through an as yet undetermined distribution function  $f(c)$ . Thus we may write

$$d^3N = f(c)dc d\Omega = f(c)dc \sin \theta d\theta d\phi. \quad . . . . . (4)$$

The total number of molecules in a cubic centimetre, given by integrating over the unit sphere and over all speeds, is equal to

$$N = 4\pi \int f(c)dc. \quad . . . . . (4')$$

The number of molecules in  $d\Omega$  hitting unit area of the wall in time  $\Delta t$  is the same as the number contained in a cylinder whose base has unit area and whose length

\* The notation  $d^3N$  is used to indicate that we are dealing with a third-order differential.

The assumptions made here are not hard to understand in terms of an analogy. If one asks in a large assemblage of people, “Who among those present is exactly forty-five years old this minute?” it is almost certain that no one will respond. But if one asks instead, “Who is between forty-five and forty-six years old?” several individuals may make themselves known, and the greater the interval the greater the number of expected responses.

is  $c\Delta t$ . Since the axis of this cylinder makes an angle  $\theta$  with the normal to the wall, the number is  $d^3Nc \cdot \cos\theta \cdot \Delta t$ . Each molecule exerts a force  $2mc \cos\theta/\Delta t$  on collision, so that by integrating over all directions and speeds the pressure is found to be

$$p = \int_0^\infty \int_0^{\pi/2} \int_0^{2\pi} f(c) dc \sin\theta \cdot 2mc^2 \cos^2\theta d\theta d\phi = \frac{4\pi}{3} m \int_0^\infty f(c) c^2 dc. \quad (5)$$

If we deal with any multiplying function  $\psi$  of  $c$ , it is logical to define the mean value  $\bar{\psi}$  as

$$\bar{\psi} = \frac{\int_0^\infty \psi(c) f(c) dc}{\int_0^\infty f(c) dc} = \frac{4\pi}{N} \int_0^\infty \psi(c) f(c) dc. \quad (6)$$

Hence, (5) becomes

$$p = \frac{Nm\bar{c}^2}{3}. \quad (7)$$

Consider, now, one mol of gas. This contains  $L$  molecules, so that the number of molecules in one cubic centimetre is  $N = L/v$ . If we substitute this value in Bernoulli's equation, we obtain

$$pv = \frac{Lm\bar{c}^2}{3}. \quad (8)$$

This is the law of Boyle and Mariotte. Identifying the right member of (8) with the right member of the gas equation (p. 495), we obtain a relation between temperature and mean square molecular velocity:

$$RT = \frac{Lm\bar{c}^2}{3} \quad \text{or} \quad \bar{c}^2 = \frac{3RT}{Lm}. \quad (9)$$

This shows that the mean square velocity of the molecules is directly proportional to the absolute temperature and inversely proportional to the molecular weight  $Lm$ .

A complete kinetic theory derivation of the entire equation of state requires statistical methods, and will be carried out in the following chapter.

## 2. Number of Collisions and Mean Free Path for Real Gases.

The assumption that the molecules are particles with no extension in space is of course a drastic simplification of the actual situation. We obtain a better representation of the facts if we attribute finite size to the molecules, considering them to be elastic spheres of radius  $r$ . The resulting correction of the equation of an ideal gas has been considered in detail on p. 497. At this time we wish to consider only the collisions between the moving molecules. If all the molecules



with the exception of a single one were at rest, it would be a very simple matter to calculate the average number of collisions experienced by this molecule. This computation will be carried out here. There is a collision whenever the distance between the centre of the selected molecule and the centre of any other molecule is less than  $d = 2r$ . The sphere of radius  $d$ , into which the centre of no other molecule can penetrate, is called the "sphere of influence" of the molecule. If the average velocity of the selected molecule is  $\bar{c}$ , the sphere of influence sweeps out a cylinder of volume  $\pi d^2 \bar{c}$  in one second. Let there be  $N$  molecules per cubic centimetre; this cylinder will then contain  $N\pi d^2 \bar{c}$  molecules and, under the assumptions made, there will be

$$Z' = \pi N d^2 \bar{c} \quad . \quad . \quad . \quad . \quad . \quad . \quad (10)$$

collisions per second.

In reality, the other molecules are also in motion, and so the mean velocity of a single molecule must be replaced by the *mean relative velocity*  $\bar{c}_r$  of the molecules with respect to each other. This velocity is readily computed if, for simplicity, we neglect the distribution of velocities and calculate with a single velocity  $\bar{c}$ . But since all possible directions between velocity vectors occur, we still obtain a distribution of relative velocities, the mean value of which we shall calculate. If  $\theta$  is the angle between the velocity vectors of two molecules, the relative velocity is

$$c_r = \sqrt{c^2 + c^2 - 2c^2 \cos \theta} = 2\bar{c} \sin \frac{\theta}{2}.$$

The spatial average of the trigonometric function  $\sin(\theta/2)$  must be determined. In general, the mean of a function of the spherical coordinates is obtained by multiplying the function by the element of area of the unit sphere, integrating over the sphere, and dividing by the area of the sphere, i.e. by  $4\pi$ :

$$\bar{f}(\theta, \phi) = \frac{1}{4\pi} \int_0^\pi \int_0^{2\pi} f(\theta, \phi) \sin \theta \, d\theta \, d\phi.$$

In the present case we have

$$\overline{\sin \frac{\theta}{2}} = \frac{2}{4\pi} \int_0^\pi \int_0^{2\pi} \sin \frac{\theta}{2} \sin \frac{\theta}{2} \cos \frac{\theta}{2} \, d\theta \, d\phi = \frac{2}{3}.$$

Hence the number of collisions per second becomes

$$Z = \frac{4\pi}{3} N d^2 \bar{c}. \quad . \quad . \quad . \quad . \quad . \quad (11)$$

The more rigorous calculation, taking into account the distribution of velocities, leads to the formula

$$Z = \sqrt{2}\pi N d^2 \bar{c}. \quad . \quad . \quad . \quad . \quad . \quad (11')$$

The difference in the numerical factors  $4/3 = 1.33$  and  $\sqrt{2} = 1.41$  is about 6 per cent. In view of the inaccuracy inherent in the nature of the calculation, and in particular the indefiniteness of the concept of atomic sphere of influence, the difference is of no practical account.

The *mean free path*  $l$  is defined as the average distance traversed by a molecule between two collisions, i.e.

$$l = \frac{\bar{c}}{Z} = \frac{3}{4\pi N d^2}, \quad . \quad . \quad . \quad . \quad . \quad (12)$$

or, according to the more rigorous calculation,

$$l = \frac{1}{\sqrt{2}\pi N d^2}. \quad . \quad . \quad . \quad . \quad . \quad (12')$$

Let us investigate in more detail the significance of the mean free path. Take  $W(x)$  to be the probability that a molecule travel a distance  $x$  without suffering a collision. Then the probability of traversing the distance  $x + dx$  is  $W(x + dx)$ . This is composed of the probability that the path  $x$  will be covered and the probability that the added segment  $dx$  will be traversed also without collision. But according to the Theory of Probability, the probability that both events will occur is given by the product of the separate probabilities, since the two are independent. Then

$$W(x + dx) = W(x) + \frac{dW}{dx} dx = W(x) W(dx). \quad . \quad (13)$$

Now the probability of free passage through a layer of thickness  $dx$  is readily found. Imagine a cylinder whose length  $dx$  is parallel to the direction of motion and whose cross-sectional area is unity. The probability of a collision within the distance  $dx$  is obviously given by the ratio of the combined areas of the front surface, through which molecules which later suffer collisions pass, to the entire front surface of unit area. The former quantity is equal to the sum of the projections, on the front surface, of all the spheres of influence lying within the cylinder,\* i.e.  $\pi N d^2 dx$ . Hence the probability that the segment  $dx$  is traversed without collision is given by

$$W(dx) = 1 - \pi N d^2 dx. \quad . \quad . \quad . \quad . \quad . \quad (13')$$

\* This is valid only for an infinitely thin layer; otherwise, the projections of the spheres of influence will overlap.





We wish to compute the coefficient of viscosity  $\eta$  on the basis of kinetic theory concepts. According to Newton's Second Law, the force which the upper layer exerts on the lower is equal to the rate of change of momentum of the latter. But the change of momentum can be calculated on the molecular theory. As a result of the thermal agitation, individual molecules pass from the upper layer—where the flow velocity is greater—into the lower layer. These molecules possess forward momentum corresponding to that existing in the layer where they experienced their last impact. In the same way, molecules diffusing from the lower, less rapidly moving layer into the upper stratum cause a decrease of forward momentum there. The process may be likened to the case of two trains moving in the same direction along parallel tracks but with somewhat different speeds. If the passengers jump from one train to another, the faster one will be slowed down, the slower one will be accelerated.

We shall now express this idea in mathematical form. Let the number of molecules per unit volume having the velocity components  $v_x$  and  $v_z$  be given by  $\nu(v_x, v_z)$ , where  $v$  is the resultant of the gas-kinetic and flow velocities. Then the number of molecules passing through unit area of the  $xy$ -plane per second is  $\nu v_z$ ; these molecules transport momentum whose  $x$ -component amounts to  $\nu(v_x, v_z)v_z m v_x$ . In all, the force on each square centimetre is

$$\frac{F}{S} = \sum \nu(v_x, v_z) v_z m v_x, \quad . . . . . (17)$$

where the summation is to extend over all values of  $v_x$  and  $v_z$  which occur. The evaluation of this sum is accomplished as follows. First we have

$$v_x = c_x + u, \quad v_z = c_z. \quad . . . . . (18)$$

If a molecule has experienced its last collision in the layer  $z$ ,

$$u(z) = u_0 + z \frac{du}{dz}, \quad . . . . . (19)$$

where  $u_0$  is the velocity in the stratum  $z = 0$ . Using this value, we obtain

$$\sum \nu(v_x, v_z) m v_x v_z = \sum \nu(v_x, v_z) m \left( c_x + u_0 + z \frac{du}{dz} \right) c_z. \quad (20)$$

The terms  $\sum \nu c_x c_z$  and  $\sum \nu u_0 c_z$  cancel out, since there are just as many molecules with positive  $c_x$  and  $c_z$  as with negative. However, the sum  $\sum \nu(v_x, v_z) z c_z$  does not cancel out, as may be seen by putting  $z = r \cos \theta$ ,

$c_z = c \cos \theta$ , in which case the factor  $\cos^2 \theta$  appears. If we replace  $r$ ,  $c$  and  $\cos^2 \theta$  in

$$\sum \nu(v_x, v_z) m v_x v_z = \sum \nu(v_x, v_z) m r c \cos^2 \theta \frac{du}{dz} \quad . \quad . \quad (21)$$

by their mean values  $l$ ,\*  $\bar{c}$  and  $1/3$  respectively, we obtain, on account of  $\sum \nu(v_x, v_z) = N$ ,

$$\frac{F}{S} = \frac{N m l \bar{c}}{3} \frac{du}{dz}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (22)$$

whence, by comparison with (16),

$$\eta = \frac{N m l \bar{c}}{3}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (23)$$

Inserting the value of the mean free path from equation (12) (p. 567),

$$\eta = \frac{m \bar{c}}{4 \pi d^2} = \frac{M \bar{c}}{4 \pi L d^2}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (24)$$

where  $M$  is the molecular weight of the gas and  $L$  is Avogadro's Number. This formula is of importance on account of the fact that  $N$ , the number of molecules per cubic centimetre, does not appear in it. This means that the viscosity is independent of the density and hence of the pressure of the gas. Although this result may seem contrary to intuitive views, it is well substantiated by experiment. It is only at very low pressures, where the mean free path becomes of the order of magnitude of the dimensions of the container, that the theory must be modified. In this case there is a dependence on density. Qualitatively, it is easy to see why the viscosity is independent of the density under normal conditions: despite the fact that the number of momentum carriers increases with the pressure, their mutual hindrance also increases in the same direction, thus tending to offset any change.

### (b) Conduction of heat at moderate pressures

Instead of a velocity gradient, assume that there is now a temperature gradient in the gas in the direction of the  $z$ -axis. According to p. 489, the flow of heat through the  $xy$ -plane is given by

$$-w = k \frac{dT}{dz}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (25)$$

\* At first sight, it may seem questionable whether the mean free path  $l$  may be substituted for the average value of  $r$ , since the next collisions may not take place precisely in the plane  $z = 0$ . In this connexion it must be remembered that there is no propagation of probability for the collisions; i.e. regardless of which point of the path we reckon from, the probability of traversing the segment  $r$  is  $e^{-r/l}$ , so that  $r = l$ . Hence the mean value of all paths described by the molecules after passing the plane  $z = 0$  is  $l$ , and the same is true, of course, for the original question.

where the temperature is denoted by  $T$ . We can calculate the coefficient of heat conduction on the basis of the kinetic theory, just as we did for viscosity. Instead of a transport of momentum, the passage of molecules from one layer to another effects a transport of thermal energy. Again, each molecule is assumed to possess the energy of thermal agitation corresponding to the layer in which it experienced its last collision. The thermal energy per molecule is

$$u = \frac{c_v}{L} T = \frac{c_v}{L} \left( T_0 + \frac{dT}{dz} z \right), \quad \cdot \cdot \cdot \cdot (26)$$

where  $c_v$  is the specific heat. If there are  $\nu(u, c_z)$  molecules per cubic centimetre with the velocity component  $c_z$  and thermal energy  $u$ , the flow of heat is given by

$$-w = \Sigma \nu(u, c_z) c_z \frac{c_v}{L} \left( T_0 + \frac{dT}{dz} z \right) = \Sigma \nu(u, c_z) \cdot \frac{c c_v r}{L} \cos^2 \theta \frac{dT}{dz}. \quad (27)$$

Replacing  $c$ ,  $r$  and  $\cos^2 \theta$  by their average values, we obtain

$$\frac{N \bar{c} c_v}{3L} \frac{dT}{dz} = \frac{c_v \bar{c}}{4\pi L d^2} \frac{dT}{dz}; \quad \cdot \cdot \cdot \cdot (28)$$

$$\text{i.e.} \quad k = \frac{c_v \bar{c}}{4\pi L d^2}. \quad \cdot \cdot \cdot \cdot (29)$$

Thus the coefficient of thermal conduction is also independent of the pressure.

It is, of course, not quite correct to replace the quantities  $c$ ,  $r$  and  $\cos^2 \theta$  by their mean values in the sum, but again the rigorous theory yields a numerical factor which differs but little from that obtained here. The resulting factor is 0.247, which replaces  $1/4$ .

### (c) *Viscosity and heat conduction at low pressures*

Consider next the important case of very low pressures, where the free path is large compared with the dimensions of the vessel and so essentially only collisions with the wall take place.\*

Imagine the gas placed between two parallel walls, one of which moves past the other with a speed  $u$ . The moving plate exerts a force on the one at rest, since the gas molecules that hit the moving plate have its forward speed added to any velocity they already possess, and communicate this to the other plate at their next impact with it. In order to compute the momentum communicated, consider the molecules whose directions of motion lie in a zone forming the angles  $\theta$  and  $\theta + d\theta$  with the normal to the wall. Since we are concerned only with

\* In order to gain an idea of the numerical magnitude, notice that at room temperature the vapour pressure of mercury is  $10^{-3}$  mm. and that the mean free path in this instance is about 10 cm.



those molecules that move *toward* the wall, we must take half the number. The number in each cubic centimetre will then be

$$dN = C \sin \theta d\theta,$$

where the proportionality factor  $C$ , evaluated by integration over a hemisphere, turns out to be  $N/2$ . If the mean speed is  $\bar{c}$ , the number of impacts per second per square centimetre becomes

$$Z = \frac{N\bar{c}}{2} \int_0^{\pi/2} \sin \theta \cos \theta d\theta = \frac{N\bar{c}}{4}.$$

Since each molecule gets an additional speed  $u$ , the momentum transfer per second, and hence the viscous force per unit area, amounts to

$$F = \frac{N\bar{c}}{4} um = \frac{\rho\bar{c}}{4} u, \quad . . . . . (30)$$

where  $\rho$  is the density and is equal to  $pM/RT$ . Then, by (9), p. 565, we have approximately

$$F = \frac{p}{4} \sqrt{\frac{3M}{RT}} u. \quad . . . . . (30')$$

Inasmuch as there is no velocity gradient in the space occupied by the gas itself, we cannot speak of a *coefficient* in the sense of equation (16); rather, the momentum transfer is independent of the distance apart of the walls.

In the same way, one finds that the heat transfer per unit area at very low pressures amounts to

$$W = \frac{3}{8} p \sqrt{\frac{3R}{MT}} (T_1 - T_2), \quad . . . . . (31)$$

where  $T$  is to be set equal to  $\frac{1}{2}(T_1 + T_2)$ . As long as the mean free path is large compared with the dimensions of the vessel, viscosity and heat conductivity are proportional to the pressure but independent of the dimensions of the container.

(d) *Kinetic Theory determination of Avogadro's Number  
and of molecular diameter*

Since we may replace  $\bar{c}$  approximately by  $\sqrt{\bar{c}^2} = \sqrt{3RT/M}$  (the exact relation, as given on p. 592, is  $\bar{c} = \sqrt{8/3\pi} \cdot \sqrt{\bar{c}^2}$ ), the only unknowns remaining in equations (24) and (29) are  $L$  and  $d$ . Since both occur only in the combination  $Ld^2$ , a further relation is needed to determine these quantities separately. Such an equation is that for

the van der Waals  $b$  which, as we already saw on p. 497, is equal to four times the space occupied by the spherical molecules:

$$b = 4 \cdot \frac{4\pi}{3} d^3 L. \quad \dots \dots \dots (32)$$

Solution of the two equations for  $L$  and  $d$  yields numbers of the order of  $10^{-8}$  cm. for the molecular diameters and about  $6 \times 10^{23}$  for  $L$ , which was first determined in this way by Loschmidt. However, the accuracy of this kinetic theory determination of  $L$  is not as great as that involving the electronic charge  $e = F/L$ , from which is obtained the generally accepted value  $6.02 \times 10^{23}$ . With reference to the gas-kinetic diameter, it must be remembered that this is not the actual diameter of a small, impenetrable sphere, but rather—according to the way it was defined—the radius of the sphere of influence for molecular collisions. In view of the electrical structure of atoms and molecules, this sphere of influence is a result of the mutual electrical influences of the colliding bodies. This does not mean that other methods of determining atomic dimensions would necessarily yield the same result. We shall find that the linear dimensions of the space occupied by the electrons is of the order of magnitude of  $10^{-8}$  cm. All experiments which depend on the electron shells lead to dimensions of this magnitude, but experiments on the passage of  $\alpha$ -particles through matter lead to values of the order of  $10^{-12}$  cm. for the diameter of the “centre”, i.e. the nucleus of the atom.

#### 4. Derivation of some Properties of Crystals on the Molecular Theory.

##### (a) *Density and lattice constant*

While the beginnings of a molecular theory of liquids have been obtained by starting with a van der Waals gas and assuming the cohesive forces to be very large, there is no complete theory which explains all the main properties, as there is for gases. However, for crystals—which we may regard as the normal form of the solid state—the theory has been developed very extensively. On account of the elaborate mathematical equipment needed, we can give only the simplest parts of the theory. M. von Laue's discovery of X-ray interference in crystals (p. 385) proved the existence of the lattice structure which mineralogists had long assumed. To-day a large number of lattice types are known. We shall here consider only the simplest case—that of sodium chloride. Measure-

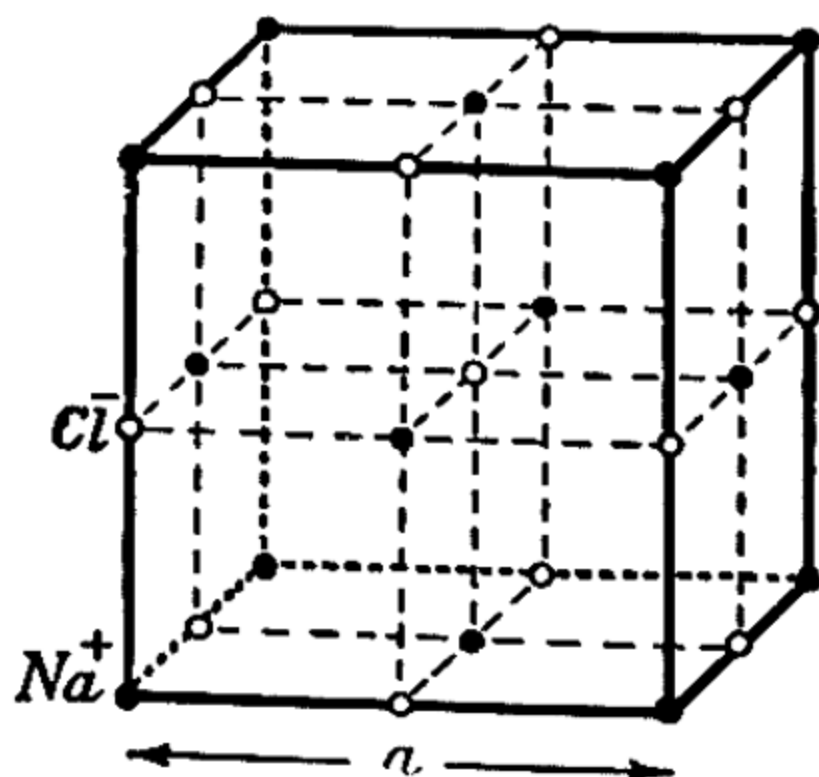


Fig. 1

ment of the intensity of X-rays reflected from different lattice planes shows that the corners of this lattice are occupied not by the neutral atoms as is the case for diamond, for example—but by the same  $\text{Na}^+$  and  $\text{Cl}^-$  ions as are encountered in electrolysis. The ions of each kind are distributed upon a cubic lattice, the two lattices being interspaced, so that along any line the ions of the two varieties alternate (“face-centred” lattice). If no distinction is made between the two kinds of ion, the crystal represents a simple cubic lattice, in which only the corners of the cubes are occupied (cf. fig. 1). The *lattice constant*  $a$  is the distance between two homologous ions of the same kind. An *elementary cell* is the smallest unit by whose translation the entire lattice can be generated. In the present case it is a cube of edge  $a$ . In order to calculate the density all that is necessary is to count the number of Na and Cl ions in the elementary cell. It must be borne in mind, however, that each corner ion belongs to eight cubes, and so is to be counted only as  $1/8$ ; similarly, a point on an edge is to be counted as  $1/4$ , one on a bounding face,  $1/2$ . Thus there are  $(8/8 + 6/2)$   $\text{Na}^+$  ions and  $(12/4 + 1/1)$   $\text{Cl}^-$  ions corresponding to each elementary cell. Representing the masses of the individual atoms by  $M_{\text{Na}}/L$  and  $M_{\text{Cl}}/L$  respectively, the density is

$$\rho = \frac{4(M_{\text{Na}} + M_{\text{Cl}})}{La^3} = \frac{M}{\frac{La^3}{4}} = \frac{M}{v} \quad \dots \quad (33)$$

Hence the lattice constant is given by

$$a = \sqrt[3]{\frac{4(M_{\text{Na}} + M_{\text{Cl}})}{L\rho}} \quad \dots \quad (34)$$

This quantity is used to measure the wave-length of X-rays. The absolute value is again limited by the accuracy to which the Avogadro constant is known. Recently it has been found possible to use ruled gratings, of known grating constant, for X-rays, thus measuring the wave-length directly. The value of the wave-length may then be used to determine the lattice constant; this, together with the density, leads to a new determination of  $L$ . The value thus found agrees admirably with that obtained from recent oil-drop determinations of  $e$ .

The *method* given here for obtaining an equation connecting the density and the lattice constant is, of course, perfectly general although equation (33) is not.

(b) *Nature of the cohesive forces in a crystal. Lattice energy of polar crystals*

What holds a crystal lattice together? What forces hinder a further decrease in volume, so that even very great pressures succeed in pro-



ducing only a slight contraction of a solid body? The first question is easy to answer for crystals which are built up of ions, the so-called polar crystals: the electrostatic attraction between the ions holds the structure together. The second question corresponds to that concerning the nature of the interatomic or intermolecular forces which give rise to the spheres of influence. Instead of regarding the ions as spheres of a definite radius, it is more natural to introduce forces of repulsion varying as a high inverse power of the distance. The

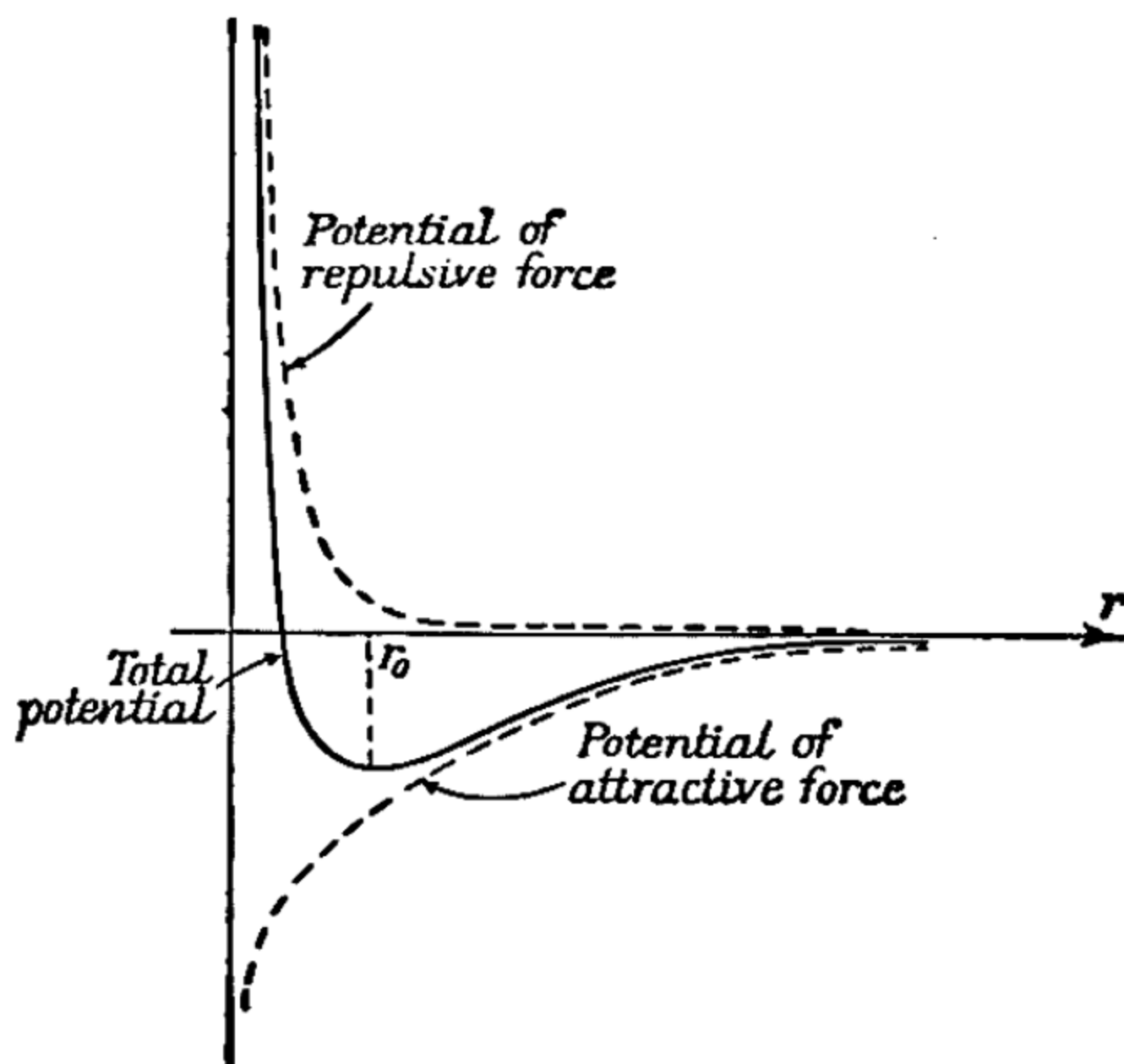


Fig. 2

higher the power, the greater the relative effect at small distances, and the better the correspondence to the collision of two rigid spheres, where the large forces which come into play at the instant of impact prevent a closer approach. The Quantum Theory has made clear the nature of the interatomic forces: they must arise when the electron shells interpenetrate. For the alkali halides the inverse tenth power of the separation proves to be the appropriate law of force. In fact, the Quantum Theory shows that a single power is really not sufficient—the exponent itself depends upon the distance.

The Quantum Theory has also thrown light on the nature of the forces of attraction in non-polar crystals, like diamond. These forces correspond to the binding forces between uncharged atoms, e.g. the  $H$  atoms in the hydrogen molecule, and are explainable on the quantum basis (cf. p. 712).

If there were but two ions present, the mutual potential energy

of an Na ion and a Cl ion, say, would be given by

$$u = -\frac{e^2}{r} + \frac{\beta}{r^9},$$

where  $\beta$  is a factor not yet determined. It would be incorrect, however, to say that the potential energy of the entire lattice is simply the above energy  $u$  multiplied by the number of atom pairs present, for since the electrostatic potential decreases only with  $1/r$ , the more remote ions also contribute to the potential at any point. The somewhat troublesome computation of the potential due to alternate positive and negative charges was made by Madelung, but cannot be reproduced here. The result is that the potential energy per ion bond is  $-0.2905(e^2/r)$ . We therefore put

$$u = -0.2905 \frac{e^2}{r} + \frac{\beta}{r^9}. \quad . \quad . \quad . \quad . \quad (35)$$

If we make a graph of this potential energy function in terms of the distance  $r$ , we obtain a curve having a minimum, since the negative term preponderates for  $r$  large, the positive term for  $r$  small (fig. 2). According to the laws of mechanics, this minimum corresponds to the equilibrium distance  $r_0$ . But the minimum condition  $du/dr = 0$  can be interpreted also to mean that the force—which is obtained by taking the potential gradient—vanishes at the equilibrium distance; i.e. the forces of attraction and repulsion are of equal magnitude. This serves to determine  $\beta$ , for

$$-0.2905 \frac{e^2}{r_0^2} + \frac{9\beta}{r_0^{10}} = 0,$$

whence 
$$\beta = \frac{1}{9} (0.2905) e^2 r_0^8, \quad . \quad . \quad . \quad . \quad (36)$$

and thus, in general,

$$u = -0.2905 e^2 \left( \frac{1}{r} - \frac{1}{9} \frac{r_0^8}{r^9} \right). \quad . \quad . \quad . \quad . \quad (37)$$

We obtain the total potential energy of the lattice in the position of equilibrium ( $r = r_0$ ) in the following way. In the NaCl lattice, six bonds connect each ion with the neighbouring ions of opposite sign. Since we count each bond but once, we may consider only the ions of one sign, say the  $\text{Na}^+$  ions. The lattice contains  $L$   $\text{Na}^+$  ions per mol, so that the energy of each pair must be multiplied by  $6L$ . We then obtain the lattice energy per mol as

$$u_l = -1.743 L e^2 \frac{8}{9 r_0}. \quad . \quad . \quad . \quad . \quad (38)$$

Using the fact that  $r_0 = a/2$  and employing equation (34) (p. 574), we have

$$\left. \begin{aligned} u_i &= -547 \times 10^3 \sqrt[3]{\frac{\rho}{M_{\text{Na}} + M_{\text{Cl}}}}; \\ u_i &= -547 \times 10^3 \sqrt[3]{\frac{\rho}{M}}, \end{aligned} \right\} \text{ in general,} \quad \dots \quad (39)$$

where the energy is expressed in calories.\* This lattice energy, valid for all salts of the NaCl type, is negative, since energy is liberated when the oppositely charged ions approach each other. This is then the work which must be done to separate the parts of the lattice and scatter them infinitely far apart. It is not equal to the heat of vaporization, for in that case the lattice is converted into molecules.

(c) *The compressibility of polar crystals*

The potential curve shows why a crystal resists changes in volume. As soon as the distance between two ions is increased from its equilibrium value, a large increase in potential energy occurs, so that a large amount of work must be expended. The force per ion pair is

$$F = -0.2905 \frac{e^2}{r^2} + 0.2905 \frac{e^2 r_0^8}{r^{10}}. \quad \dots \quad (40)$$

This force vanishes in the equilibrium position. We develop  $F$  in a Taylor's Series at the point  $r_0$  and obtain

$$\begin{aligned} F &= 0 + \frac{dF}{dr} (r - r_0) + \dots = 0.2905e^2 \left( \frac{2}{r_0^3} - \frac{10r_0^8}{r_0^{11}} \right) (r - r_0) \dots \\ &= -\frac{2.324e^2}{r_0^3} (r - r_0) \dots \quad (41) \end{aligned}$$

The work done by the internal forces when one mol of crystal is so compressed that all interionic distances are changed from  $r_0$  to  $r$  is

$$\begin{aligned} W &= 6L \int_{r_0}^r F dr = -6 \times 2.324 \frac{Le^2}{r_0^3} \int_{r_0}^r (r - r_0) dr \\ &= -13.94 \frac{Le^2}{r_0^3} \frac{(r - r_0)^2}{2}. \quad \dots \quad (42) \end{aligned}$$

The negative value shows that work must be supplied from without. But the work done by the internal forces is given by the integral

$$\int_{v_0}^v p dv.$$

\* The lattice energy is usually expressed in cal/mol, since it represents a chemical heat of reaction.



Since the pressure is zero at the start and  $P$  at the end, we obtain an approximation to the integral by multiplying the total change in volume by the mean pressure  $P/2$ . But according to equation (33) (p. 574) the change in volume is given by

$$\Delta v = \Delta \left( \frac{La^3}{4} \right) = \Delta (2Lr_0^3) = 6Lr_0^2(r - r_0). \quad (43)$$

Equating the two expressions for the work,

$$- \frac{13.94Le}{2r_0^3} (r - r_0)^2 = 6Lr_0^2(r - r_0) \frac{P}{2}, \quad (44)$$

from which

$$P = - \frac{13.94e^2(r - r_0)}{6r_0^5} = - \frac{13.94e^2\Delta v}{36Lr_0^5 \cdot r_0^2} = - \frac{13.94e^2\Delta v}{18r_0^4v} = - \frac{1}{\kappa} \frac{\Delta v}{v}$$

and the compressibility  $\kappa$  becomes

$$\kappa = \frac{18r_0^4}{13.94e^2} = 5.60 \times 10^{18}r_0^4 = 3.50 \times 10^{17}a^4. \quad (45)$$

This relation is very well substantiated experimentally, but it should be noted that the inverse tenth power, which has no other basis at present, is fitted to the compressibility.

#### (d) Qualitative explanation of the thermal expansion of crystals

The thermal motions of the particles constituting a crystal can consist only of vibrations of these particles about the equilibrium positions to which they are bound. When an atom is in the position of equilibrium it is subject to no force, but as soon as it leaves this position, restoring forces arise which, by equation (41), are proportional to the displacement  $r - r_0$ , and so give rise to harmonic vibrations. This proportionality between force and elongation, however, represents only a very rough first approximation. If we consider the potential curve (fig. 2, p. 575), we see that for larger displacements, where the higher terms of the power series must be taken into account, the forces become unsymmetric, since the curve is much steeper on the approaching side than on the receding side. Hence if an atom has a given amount of kinetic energy when it passes through the position of equilibrium, this will become equal to the potential energy at a smaller distance from the rest position on the near side than on the far side. This means that the reversal of motion will take place sooner on the near side. On the average, then, the distance between two atoms will be increased, and this is the reason for the thermal expansion of crystals (cf. *Ex.* 40, p. 103).

## CHAPTER XXXV

# THE CLASSICAL STATISTICS OF BOLTZMANN

## 1. Entropy and Probability.

We wish now to treat molecular phenomena somewhat more abstractly by starting from a perfectly general relation which bridges the gap between thermodynamics and statistics, and deducing from it a large number of consequences.

From the example of the diffusion of two gases through each other—a process involving an increase of entropy—we recognize that there must be an intimate connexion between the entropy of a system in a given state and the probability of finding it in this state. If an urn containing equal numbers of white and red balls is thoroughly shaken, the probability of finding a practically uniform mixture is enormously greater than that of finding only white balls in the upper half and only red balls in the lower. Thus the probability has increased to a large extent in the process of mixing; this is equally true for the diffusion of two gases, for which this model is intended. Is it possible to formulate this relationship mathematically? The entropy of two *completely independent* systems—e.g. 1 cm.<sup>3</sup> of oxygen at standard conditions in London and 1 m.<sup>3</sup> of hydrogen at a pressure of 10 atm. and a temperature of 300° abs in Manchester—is, according to the definition of entropy as an additive quantity, equal to the sum of the separate entropies, if we consider the two systems as a *whole*. The probability of finding both systems at the respective given conditions is, by the law of probability for independent events, equal to the product of the separate probabilities. We thus have

$$S_{12} = S_1 + S_2, \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad (1)$$

$$W_{12} = W_1 W_2. \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad (2)$$

What function satisfies the condition that

$$f(x_1 x_2) = f(x_1) + f(x_2)? \quad . \quad . \quad . \quad . \quad . \quad (3)$$

Obviously the function  $k \log x$  is appropriate. We therefore take

$$S = k \log W. \quad \cdot \cdot \cdot \cdot \cdot \cdot (4)$$

This is the celebrated Boltzmann Relation—one of the most important equations in physics. We shall find later that the factor of proportionality  $k$  is the “gas constant per molecule”  $R/L$ .

How is the probability of occurrence of a thermodynamical state to be expressed? The probability of an event is usually defined in the Theory of Probability as the ratio of the number of cases favourable to the occurrence of the event to the total number of possible cases. This probability has a maximum value of unity, corresponding to *certainty*. In thermodynamics, however, it has been found more convenient to take simply the number of favourable cases as the measure of thermodynamic probability. Thus the thermodynamic probability is a large number.

In counting the number of cases favourable to a given thermodynamical state, the expression  $N!$  (“ $N$  factorial” =  $N(N-1)(N-2)\dots 1$ ) occurs, as in other branches of combinatorial analysis. Since we deal here with numbers of the order of magnitude of  $L$ , i.e.  $10^{23}$ , the evaluation of  $N!$  would be out of the question.

Fortunately, however, there is an approximate formula due to Stirling which gives a good approximation when  $N$  is large. If we write

$$\log N! = \log 1 + \log 2 + \log 3 + \log 4 + \dots \log N,$$

we may look upon the series as a sum of rectangular strips whose breadth is  $\Delta x = 1$  and whose height is  $\log x$ . When  $x$  is large the intervals of unit width may be regarded as differentials, so that the sum may be replaced by the integral

$$\int_1^N \log x dx = [x \log x - x]_1^N = N \log N - N + 1.$$

Neglecting 1 as compared with  $N$  and taking the antilogarithm of both members, we have the useful approximate formula for large values of  $N$ :\*

$$N! \approx \left(\frac{N}{e}\right)^N \cdot \dots \cdot \dots \cdot \dots \cdot \dots \quad (5)$$

## 2. Calculation of the most Probable Distribution of Density in an Ideal Gas.

We shall calculate the thermodynamic probability and its maximum value, corresponding to the state of equilibrium, for the density distribution in an ideal gas. Here we already know the result: at equilibrium the density must, of course, be the same at all points. Although the case is trivial, it is—on account of its simplicity—suitable as an explanatory example of the method of computing probabilities.

\* Stirling's Theorem states more exactly that  $\log(N!) \approx (N + \frac{1}{2}) \log N - N + \frac{1}{2} \log(2\pi)$ , the error being approximately  $1/(12N)$ .



In the thermodynamic, macroscopic sense we characterize a given state by giving the density as a function of position. But for an observer capable of seeing and distinguishing between individual molecules, a given state would be defined only by stating the co-ordinates of each molecule. It is clear that any particular "macro-state" can be realized by a large number of different "micro-states", for the interchange of two molecules does not alter the density distribution in the least. *We therefore look upon the number of micro-distributions corresponding to a given macro-state as a measure of the probability of occurrence of the thermodynamic state.* In order to compute this number, we divide the volume into a number of equal cells of volume  $\Delta\tau$ . The density is then given by the ratio of the number of molecules  $N_i$  in each cell to the volume of the cell. The cells are to be small, but nevertheless large enough to include a great number of molecules, so that the density  $N_i/\Delta\tau$  remains a continuous function of position. If we were to make each cell of the order of magnitude of the average space available to a molecule, we would notice relatively large discontinuities in density from cell to cell, according as 0, 1, 2, 3 . . . molecules happen to be in each.

Let the cells and the molecules be numbered, and suppose that the molecules are thrown into the subdivided volume much like a handful of glass beads. Since the cells are all of equal size, it is equally probable whether a molecule  $X$  falls into the cell  $y$  or the cell  $z$ . That is, the division into equal cells has assured equal *a priori probability* of the cells, as it is usually expressed. This equal a priori probability is not disturbed if there are already several molecules in a cell; it is assumed that there is sufficient room for others.\* Thus, in order to compute the thermodynamic probability, we need only count the number of ways in which  $N$  molecules can be distributed among  $Z$  cells in such way that there are  $N_i$  molecules in the  $i$ th cell. How often this state can be produced is best seen from an example. Let the following distribution be given:

Cell:	1	2	3	4	5	6
Number of molecules:	1	0	2	0	3	4.

Such a distribution of ten molecules among six cells may be realized, for example, in this way:

Molecule number:	1	2	3	4	5	6	7	8	9	10
In cell number:	6	1	6	5	6	5	3	3	5	6.

Evidently there are as many micro-states as there are permutations of the cell numbers of the molecules. It is to be noticed that the change

\* This is exactly the point of difference between the classical statistics and a new form due to Fermi. Cf. p. 632.

in position of the molecules *within* one cell does not yield a new micro-state (or "complexion" as it was termed by Gibbs and Boltzmann), for the micro-state is determined solely by giving the cell numbers. Thus the probability of the distribution is equal to the number of permutations of the lower row of the preceding table. If all the cell numbers were different this would be  $N!$ , but since, for example, the number six occurs four times, all permutations of these four numbers among each other are omitted. On account of the fourfold occurrence of the number six we then have only the  $1/4!$  part. In general,

$$W = \frac{N!}{N_1! N_2! \dots} = \frac{N!}{\prod (N_i!)} \quad \cdot \cdot \cdot \cdot (6)$$

Applying the Stirling approximation and remembering that  $\sum N_i = N$  we have

$$W = \frac{N^N}{\prod N_i^{N_i}} \quad \cdot \cdot \cdot \cdot (6')$$

We now introduce  $f(i)$ , a density function of the cell index, defined by

$$N_i = f(i) \Delta \tau_i = f(i) \Delta \tau. \quad \cdot \cdot \cdot \cdot (7)$$

As long as all the cells have the same volume, we can omit the subscript on  $\Delta \tau$ . The entropy then becomes

$$S = k(N \log N - \sum N_i \log N_i) = k(N \log N - \sum f(i) \Delta \tau \log f(i) - \sum f(i) \Delta \tau \log \Delta \tau). \quad \cdot \cdot \cdot (8)$$

The equilibrium value is characterized by the vanishing of the first variation:

$$-\delta S = 0 = \sum \delta f(i) \Delta \tau \log f(i) + \sum \delta f(i) \Delta \tau + \sum \delta f(i) \Delta \tau \log \Delta \tau. \quad (9)$$

It is to be remembered that an auxiliary condition is given by the fact that the total number of molecules is to remain constant:

$$\sum f(i) \Delta \tau = N, \quad \text{i.e.} \quad \sum \delta f(i) \Delta \tau = 0. \quad \cdot \cdot \cdot (10)$$

The variations of the distribution function are therefore not arbitrary; equation (10) must always be satisfied. If we apply the method of undetermined coefficients (p. 115), multiplying equation (10) by  $\alpha$  and adding it to equation (9), the coefficients of the variations may be set equal to zero separately, and there results

$$\log f(i) + \alpha + 1 + \log \Delta \tau = 0 \quad \text{or} \quad f(i) = \text{const.} \quad (11)$$

As expected, the most probable distribution is that of uniform density throughout. From equation (10), the value of the constant turns out to be  $N/V$ , where  $V$  is the total volume.

The division into regions of equal a priori probability is too narrow

a limitation. In many instances, e.g. in the distribution over the surface of a sphere, a division into zones of equal size is quite impossible; moreover, in quantum statistics it often happens that the individual cells must be given different a priori probabilities (*weights*), which is naturally equivalent to making the sizes of the cells different.\* How are we to proceed in this more general case in which each cell has a volume  $\Delta\tau_i$  of a priori probability  $g_i$ ? In any case, we set the total weight  $G_i$  equal to the cell volume multiplied by  $g_i$ . If the distribution is such that one of the cells has twice the volume of any of the others (which have equal volumes), the probability that a molecule  $a$  should fall in this cell will be doubled, as will that for molecules  $b$  and  $c$ . Then the probability that all three molecules fall in this cell is the product of the individual values, and so increases to  $2^3$ . In general, the factor to be applied to a combination in which  $N_i$  molecules fall in a cell having a weight  $G_i$  is  $G_i^{N_i}$ . Thus a combination in which there are  $N_1$  molecules in the first cell,  $N_2$  in the second, &c., has a factor  $G_1^{N_1} G_2^{N_2} \dots$ , and we obtain

$$W = \frac{N! \prod G_i^{N_i}}{\prod N_i!} = \frac{N^N \prod G_i^{N_i}}{\prod N_i^{N_i}} = \frac{N^N \prod G_i^{f(i) \Delta\tau_i}}{\prod [f(i) \Delta\tau_i]^{f(i) \Delta\tau_i}}. \quad (12)$$

$$S = k[N \log N + \sum f(i) \Delta\tau_i \log G_i - \sum f(i) \Delta\tau_i \log f(i) - \sum f(i) \Delta\tau_i \log \Delta\tau_i]. \quad (13)$$

The calculation of the most probable distribution is carried out exactly as in the previous example. We have immediately

$$\frac{\delta S}{k} = 0 = \sum \delta f(i) \Delta\tau_i \log G_i - \sum \delta f(i) \Delta\tau_i \log f(i) - \sum \delta f(i) \Delta\tau_i \log \Delta\tau_i, \quad (14)$$

with the auxiliary condition

$$\delta N = \sum \delta f(i) \Delta\tau_i = 0. \quad (15)$$

Multiplying by  $-a$  and adding equation (14), we obtain

$$\log G_i - \log f(i) - 1 - \log \Delta\tau_i - a = 0, \quad (16)$$

or 
$$f(i) = \text{const.} \frac{G_i}{\Delta\tau_i}. \quad (17)$$

Since  $\sum f(i) \Delta\tau_i = N$ , the constant is  $N/\sum G_i$ , so that the most probable distribution is represented by the function

$$f(i) = \frac{NG_i}{\Delta\tau_i \sum G_i} = \frac{Ng_i}{\sum g_i \Delta\tau_i}. \quad (18)$$

\* In many instances the a priori probabilities have been obtained in such manner that the connexion with experience was realized after their introduction, so that here we are really dealing with weights deduced from experience—a posteriori probabilities—whose firmer theoretical foundation (and hence transformation into true a priori probabilities) must be left to further developments.



With equal a priori probabilities  $g_i = 1$ , the result is

$$f(i) = \frac{N}{\sum \Delta \tau_i} = \frac{N}{V} \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad (19)$$

Thus even if the cells differ in size, as long as they otherwise have no special a priori probabilities, the most probable distribution is that of uniform density.

*Ex. 128.* It follows from the formulæ of p. 530. that, if a paramagnetic substance could exist in an unmagnetized state at absolute zero (this is denied by the Nernst theorem), its entropy per mol would in the simplest case be greater than in the magnetized condition by  $R \log 2$ . Deduce this result statistically, remembering that in the simplest case only the orientations parallel and anti-parallel to the field are permitted.

### 3. Representation of the Distribution of Position and Velocity by means of Phase Space. Liouville's Theorem.

We wish to know not only the position, but also the distribution of the velocities of the gas molecules in the state of equilibrium. In order to compute the probability of a given state, it is natural to seek to apply the procedure of the last section to this more complicated case. To do this, we need only replace actual three-dimensional space by a hypothetical six-dimensional space in which the co-ordinates of a point are the three space co-ordinates of a molecule together with its three velocity components.\* Thus every molecule in actual three-dimensional space has a representative point in this six-dimensional space. One might raise an objection to this procedure. In ordinary space it is immediately evident that the a priori probability of a cell is proportional to its size, but how can we be certain about the a priori probability of a cell in the hyperspace? If, instead of taking the velocity components, we choose the general momentum components  $p_k$  (cf. p. 121), the space built up of the position co-ordinates  $q_k$  and the momentum co-ordinates  $p_k$  is called a *phase space*. For this space it can be shown that the cells of equal a priori probability are of equal size or—what is the same thing—that the a priori probability of a cell of phase space is proportional to the size of this cell. For if we consider a volume element  $\Delta \tau$  of this space, representative points will enter and leave through the bounding surface of the element, and we can compute the “source-strength” (divergence) for the element exactly as in the three-dimensional case. If

$$\mathbf{r} = q_1 \mathbf{e}_1 + q_2 \mathbf{e}_2 + q_3 \mathbf{e}_3 + p_1 \mathbf{e}_4 + p_2 \mathbf{e}_5 + p_3 \mathbf{e}_6, \quad \cdot \quad (20)$$

\* For the present, we limit our considerations to monatomic molecules whose condition is uniquely determined by the position and velocity of the centre of gravity. For polyatomic molecules we have angular co-ordinates and velocities also.

the velocity  $\mathbf{C}$  of a representative point is

$$\begin{aligned}\mathbf{C} &= C_1 \mathbf{e}_1 + C_2 \mathbf{e}_2 + C_3 \mathbf{e}_3 + C_4 \mathbf{e}_4 + C_5 \mathbf{e}_5 + C_6 \mathbf{e}_6 \\ &= \frac{dq_1}{dt} \mathbf{e}_1 + \frac{dq_2}{dt} \mathbf{e}_2 + \frac{dq_3}{dt} \mathbf{e}_3 + \frac{dp_1}{dt} \mathbf{e}_4 + \frac{dp_2}{dt} \mathbf{e}_5 + \frac{dp_3}{dt} \mathbf{e}_6.\end{aligned}\quad (21)$$

Choosing a small six-dimensional cube as the element of volume, we have, exactly as in the three-dimensional case,

$$\begin{aligned}\text{div } \mathbf{C} &= \frac{\partial C_1}{\partial q_1} + \frac{\partial C_2}{\partial q_2} + \frac{\partial C_3}{\partial q_3} + \frac{\partial C_4}{\partial p_1} + \frac{\partial C_5}{\partial p_2} + \frac{\partial C_6}{\partial p_3} \\ &= \frac{\partial}{\partial q_1} \frac{dq_1}{dt} + \frac{\partial}{\partial q_2} \frac{dq_2}{dt} + \frac{\partial}{\partial q_3} \frac{dq_3}{dt} + \frac{\partial}{\partial p_1} \frac{dp_1}{dt} + \frac{\partial}{\partial p_2} \frac{dp_2}{dt} + \frac{\partial}{\partial p_3} \frac{dp_3}{dt}\end{aligned}\quad (22)$$

Now the molecules themselves obey the fundamental laws of mechanics, which we may write in the Hamiltonian form:

$$\frac{dp_k}{dt} = - \frac{\partial H}{\partial q_k}, \quad \frac{dq_k}{dt} = + \frac{\partial H}{\partial p_k}, \quad \cdot \cdot \cdot \cdot \quad (23)$$

whence it follows that

$$\frac{\partial}{\partial p_k} \frac{dp_k}{dt} + \frac{\partial}{\partial q_k} \frac{dq_k}{dt} = 0. \quad \cdot \cdot \cdot \cdot \quad (24)$$

On account of this last equation, every term in (22) cancels out the other term having the same subscript, and so the entire right-hand member vanishes. Now the vanishing of the divergence of the velocity is characteristic of the flow of an incompressible fluid; if a given number of particles (in our case, representative points) occupy an element of volume  $\Delta\tau$ , at a certain time, they will occupy another volume of equal size at a later time. Thus the relationship between two equal volume elements in the phase space is such that if one is occupied by  $N$  points the other is also filled with  $N$  points. Since, with the passing of time, all elements of the phase space are traversed by the flow, *all elements of equal size have the same a priori probability*. This is Liouville's Theorem. We have thus found a basis for computing the probability of a distribution in position *and* velocity. These considerations are not limited to a phase space of six dimensions, but hold for any hyperspace.

A diatomic molecule may be visualized by means of a model having the form of a dumb-bell. The radius of influence deduced in the elementary kinetic theory for *spherical* molecules is merely an average for various orientations of the colliding molecules. In the case of diatomic molecules we are concerned also with the orientation of the axis and the velocity of rotation about an axis normal to the *figure axis*, so that the appropriate phase space has a greater number of dimensions. But since Hamilton's Equations hold for any two canonically conjugate



variables, regardless of how many degrees of freedom the system has, the Liouville Theorem concerning the equal probability of equal elements is always valid as long as the particles whose representative points move in the phase space follow the laws of mechanics.

#### 4. The Maxwell-Boltzmann Energy Distribution.

We are now in a position to apply the method of § 2 (p. 580) literally to the phase space. By so doing, we can answer the question of how the total energy  $U$  of a system is distributed among the individual particles. The total energy of an ideal, monatomic gas, for example, is made up of the kinetic energy of translation of all the atoms. How is this energy distributed among the various atoms? Do all atoms move with the same speed, as we assumed for simplicity in the preceding chapter? A definite energy  $u(i)$  corresponds to each cell of the phase space; this is the energy possessed by a molecule whose representative point falls within this cell. For generality, assume that the cells are of various sizes. Then, by Liouville's theorem, each will have an a priori probability proportional to its size. If, as in § 2 (p. 582), we again define a distribution function  $f(i)$  by  $N_i = f(i)\Delta\tau_i$ , the probability becomes

$$W = \frac{N^N \prod \Delta\tau_i^{f(i)\Delta\tau_i}}{\prod [f(i)\Delta\tau_i]^{f(i)\Delta\tau_i}}, \quad \dots \dots \dots (25)$$

and the entropy is

$$S = kN \log N - k \sum f(i)\Delta\tau_i \log f(i). \quad \dots \dots (26)$$

The maximum value of  $S$  is to be determined under the auxiliary conditions that both the number of particles and the total energy  $U$  are to remain constant. The former condition may be stated

$$N = \sum f(i)\Delta\tau_i, \quad \text{i.e.} \quad \delta N = \sum \delta f(i)\Delta\tau_i = 0. \quad \dots (27)$$

The energy of the particles whose representative points lie in the  $i$ th cell of energy  $u(i)$  is obviously  $N_i u(i)$ , so that the total energy is

$$\sum u(i)f(i)\Delta\tau_i, \quad \text{whence} \quad \delta U = \sum \delta f(i)u(i)\Delta\tau_i = 0. \quad (28)$$

The variation of the entropy again leads to the expression

$$-\frac{\delta S}{k} = 0 = \sum \delta f(i)\Delta\tau_i \log f(i) + \sum \delta f(i)\Delta\tau_i. \quad \dots (29)$$

If we again apply the method of undetermined multipliers by multiplying the first auxiliary condition by  $\alpha$ , the second by  $\beta$ , and adding the resulting equations to (29), we obtain

$$\log f(i) + 1 + \alpha + \beta u(i) = 0, \quad \dots \dots \dots (30)$$

or setting  $e^{-(1+\alpha)} = A$ ,

$$f(i) = Ae^{-\beta u(i)}. \quad \dots \dots \dots (31)$$



The constant \*  $A$ , which is connected with the multiplier  $\alpha$ , is again determined from the total number of particles, for

$$N = A \sum e^{-\beta u(i)} \Delta \tau_i, \quad . . . . . (32)$$

so that

$$f(i) = \frac{N e^{-\beta u(i)}}{\sum e^{-\beta u(i)} \Delta \tau_i} . . . . . (33)$$

The sum appearing in the denominator is called the partition sum of the system. We shall set this extremely important quantity equal to  $\sigma$ :

$$\sigma = \sum e^{-\beta u(i)} \Delta \tau_i . . . . . (34)$$

To determine the multiplier  $\beta$  we must return to the entropy. Putting in

$$f(i) = \frac{N}{\sigma} e^{-\beta u(i)} . . . . . (33')$$

the entropy becomes

$$S = kN \log N - k \sum \frac{N}{\sigma} e^{-\beta u(i)} \Delta \tau_i (\log N - \beta u(i) - \log \sigma). \quad (35)$$

On account of (34), and since

$$\frac{N}{\sigma} \sum e^{-\beta u(i)} u(i) \Delta \tau_i = U, \quad . . . . . (36)$$

our expression simplifies to

$$S = k\beta U + kN \log \sigma. \quad . . . . . (37)$$

If  $S$  is given as a function of  $U$  and  $V$  we obtain  $T$  by differentiation, according to p. 526:

$$\frac{1}{T} = \left( \frac{\partial S}{\partial U} \right)_V . . . . . (38)$$

Further, equation (36) shows that  $U$  and  $\beta$  are functionally related so that

$$\left( \frac{\partial S}{\partial U} \right)_V = \frac{dS}{d\beta} \left( \frac{\partial \beta}{\partial U} \right)_V = \frac{dS}{d\beta} \cdot \frac{1}{\left( \frac{\partial U}{\partial \beta} \right)_V} . . . (39)$$

\* In comparing the present application of the method of multipliers with that on p. 115, where the quantities denoted there by  $\lambda$  were functions of  $x$ , one may question the constancy of  $\alpha$  and  $\beta$ . In the earlier example, the auxiliary conditions were given as functions of  $x$ , and are to be satisfied for any value of  $x$ . Here, on the other hand, two simple equations are to be satisfied—equations which no longer contain the cell subscript, since it was removed by the summation process.

Carrying out the differentiation, we have by equation (37):

$$\frac{dS}{d\beta} = kU + k\beta \frac{\partial U}{\partial \beta} + \frac{kN}{\sigma} \frac{\partial \sigma}{\partial \beta}, \quad \cdot \cdot \cdot \cdot \quad (40)$$

and from (34),  $\frac{\partial \sigma}{\partial \beta} = -\sum u(i)e^{-\beta u(i)} \Delta \tau_i = -\frac{U\sigma}{N}.$   $\cdot \cdot \cdot \cdot \quad (41)$

Hence, by (39) there remains

$$\left(\frac{\partial S}{\partial U}\right)^r = k\beta = \frac{1}{T},$$

or  $\beta = \frac{1}{kT}.$   $\cdot \cdot \cdot \cdot \cdot \cdot \cdot \cdot \quad (42)$

The distribution function then becomes

$$f(i) = \frac{Ne^{-u(i)/kT}}{\sum e^{-u(i)/kT} \Delta \tau_i}. \quad \cdot \cdot \cdot \cdot \cdot \cdot \quad (43)$$

This is the celebrated Maxwell-Boltzmann Law of Energy Distribution. In the next section we shall learn several of its applications.

Having determined  $\beta$ , we obtain a simple thermodynamical interpretation of the quantity  $\sigma$ ; putting the value  $1/kT$  for  $\beta$  in equation (37), there results

$$S = \frac{U}{T} + kN \log \sigma.$$

However, the dimensions are not yet correct in this formula. As is seen from equation (34),  $\sigma$  has the dimensions of a volume of the phase space. On the other hand, the argument of a logarithm must be dimensionless. Now, in the classical thermodynamics  $S$  is determined only to an additive constant, so that we are at liberty to add constants to the right member in order to make its dimension zero. Such a constant is  $-kN \log \Delta \tau$ , where  $\Delta \tau$  is, say, an arbitrary measure of the volume of the phase space element. Since the product of a position co-ordinate by a momentum co-ordinate has the dimensions of energy  $\times$  time, which we call action,  $[h]$ ,  $\Delta \tau$  has the dimension  $[h]^s$  if there are  $s$  position co-ordinates. If we combine this added term with the term  $kN \log \sigma$ , we see that the unit of phase space volume must be taken to be  $[h]^s$ , where the magnitude of  $[h]^s$  is still undetermined. But the Quantum Theory will show that  $[h]$  has a perfectly definite numerical value. We must therefore write

$$S = \frac{U}{T} + kN \log \frac{\sigma}{[h]^s}. \quad \cdot \cdot \cdot \cdot \cdot \cdot \quad (44)$$

Then the free energy becomes

$$F = U - TS = -kNT \log \frac{\sigma}{[h]^s}. \quad \cdot \cdot \cdot \cdot \quad (45)$$

## 5. Applications of the Maxwell-Boltzmann Energy Distribution.

(a) *Halley's Law and the equilibrium of sedimentation*

Assume that the energy of a particle of mass  $m$  depends only upon one position co-ordinate. For example, let the particle be in the earth's gravitational field and have a potential energy  $mgz$ . Then, if we take as cells the parallelepipeds of base unity and altitude  $\Delta z$ , the number of particles between the levels  $z$  and  $z + \Delta z$  is given by (43):

$$\Delta N(z) = \frac{N}{\sigma} e^{-mgz/kT} \Delta z, \quad . . . . . (46)$$

or the density is given by

$$\rho = \frac{\Delta N}{\Delta z} = \frac{N}{\sigma} e^{-mgz/kT} . . . . . (46')$$

Instead of the function  $\sigma$  we introduce here simply the density  $\rho_0$  at the level  $z = 0$ , obtaining

$$\rho = \rho_0 e^{-mgz/kT} . . . . . (47)$$

This formula includes the Halley formula (p. 192) for the variation of density (or pressure) with altitude in an isothermal atmosphere, for if we replace  $\rho_0$  in the exponent of equation (4) (p. 192) by the value  $M/v_0$ —where  $M$  is the molecular weight—and if we use the relation  $p_0 v_0 = RT$ , the exponent becomes identical with that of equation (47), since it will be shown later that  $M/R = m/k$ .\*

This formula is valid also for the distribution of visible colloid particles. If, for example, we are dealing with a mastic emulsion in water, the mass  $m$  is to be replaced by the actual mass of each particle minus the hydrostatic buoyant force. This means that in an emulsion of this kind the number of particles must decrease exponentially with the height, and this is actually observed. Thus, if the masses of individual particles can be found in any way, the counting of the numbers of particles at various levels will serve to determine the Boltzmann constant  $k$ , and hence the Avogadro constant  $L$  by means of the relation  $k = R/L$  which will be derived below.

(b) *Angular distribution of the axes of elementary magnets in a magnetic field*

It is now possible to justify the formula used on p. 455 for the distribution of the angle  $\theta$  which the axis of a dipole makes with the

\* Inasmuch as the temperature is assumed to be the same at all points, the distribution of velocities can have no effect on the distribution of particles in altitude, so that the statistical calculations for the  $z$ -co-ordinate may be performed separately.



field. The potential energy of a small magnet of moment  $p$  in a field  $H$  is given by

$$u = -pH \cos \theta. \quad . \quad . \quad . \quad . \quad . \quad (48)$$

The phase space is, in this case, the surface of a unit sphere upon which we can mark the several axial positions by points. The division into cells is best made by taking zones between the angles  $\theta$  and  $\theta + \Delta\theta$ . These zones, however, are not of equal size, since their areas, for equal intervals  $\Delta\theta$ , are proportional to  $\sin \theta \Delta\theta$ . We then have for the number of magnets whose axes make an angle with the field which lies between  $\theta$  and  $\theta + \Delta\theta$ :

$$\Delta N = \frac{2\pi N e^{pH \cos \theta / kT} \sin \theta \Delta\theta}{\sigma}. \quad . \quad . \quad . \quad . \quad (49)$$

Since there is no restriction on the size of the intervals in classical statistics, we may make them so small that the sum may be replaced by an integral. Hence in the denominator we have

$$\sigma = 2\pi \int_0^\pi e^{pH \cos \theta / kT} \sin \theta d\theta. \quad . \quad . \quad . \quad . \quad (50)$$

The rest of the computation has already been carried out on p. 456.

### (c) *Distribution of velocity in an ideal gas*

We now investigate the distribution of the velocities of the molecules of an ideal gas, limiting our considerations to the translatory motion of the molecular centres of gravity. Any rotational motions of polyatomic molecules about their centres of gravity will not be considered. These rotations can have no effect on the distribution of translational velocities, so that the results hold not only for monatomic gases, but for any molecular structures. The momenta are, in rectangular co-ordinates,

$$p_x = m\dot{x}, p_y = m\dot{y}, p_z = m\dot{z}, \text{ i.e. } p = m\sqrt{\dot{x}^2 + \dot{y}^2 + \dot{z}^2} = mc, \quad (51)$$

and we have

$$u = \frac{1}{2}mc^2 = \frac{1}{2m} (p_x^2 + p_y^2 + p_z^2) = \frac{p^2}{2m}. \quad . \quad . \quad (52)$$

The number of molecules whose centres of gravity lie between  $x$  and  $x + \Delta x$ ,  $y$  and  $y + \Delta y$ ,  $z$  and  $z + \Delta z$ , and whose momenta are between  $p_x$  and  $p_x + \Delta p_x$ ,  $p_y$  and  $p_y + \Delta p_y$ ,  $p_z$  and  $p_z + \Delta p_z$  is thus given by

$$\Delta N = \frac{N}{\sigma} e^{-(p_x^2 + p_y^2 + p_z^2)/2mkT} \Delta x \Delta y \Delta z \Delta p_x \Delta p_y \Delta p_z. \quad (53)$$

Dividing both sides by the element of volume, we obtain the number  $\Delta n$  of molecules per cubic centimetre whose momenta lie between the given limits. If we again replace the summation by an integration the denominator becomes

$$\begin{aligned}\sigma &= \iiint \iiint e^{-(p_x^2 + p_y^2 + p_z^2)/2mkT} dx dy dz dp_x dp_y dp_z \\ &= V \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{-(p_x^2 + p_y^2 + p_z^2)/2mkT} dp_x dp_y dp_z = VI^3. \quad (54)\end{aligned}$$

The integral  $I$ , which occurs here three times—with different variables, but of the same form—cannot be evaluated as an indefinite integral in terms of elementary functions, but the definite integral between the limits  $-\infty$  and  $+\infty$  can be computed with the aid of a simple artifice. Set

$$\frac{p_x}{\sqrt{2mkT}} = \xi, \quad dp_x = \sqrt{2mkT} d\xi. \quad \dots (55)$$

The integral is then

$$I = \sqrt{2mkT} \int_{-\infty}^{+\infty} e^{-\xi^2} d\xi. \quad \dots (56)$$

If the same integral is now written with the variable of integration  $\eta$  and multiplied by the integral in  $\xi$ , we obtain

$$\begin{aligned}I^2 &= 2mkT \int_{-\infty}^{+\infty} e^{-\xi^2} d\xi \int_{-\infty}^{+\infty} e^{-\eta^2} d\eta \\ &= 2mkT \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} e^{-(\xi^2 + \eta^2)} d\xi d\eta. \quad \dots (57)\end{aligned}$$

Now the double integral is readily evaluated in polar co-ordinates: If we allow the polar angle to go from 0 to  $2\pi$  and the radius vector from 0 to  $\infty$ , the plane will be completely covered, just as if the rectangular co-ordinates  $\xi$  and  $\eta$  were allowed to vary between the limits  $-\infty$  and  $+\infty$ . Hence

$$I^2 = 4\pi mkT \int_0^\infty e^{-\rho^2} \rho d\rho = -2\pi mkT [e^{-\rho^2}]_0^\infty = 2\pi mkT. \quad (58)$$

Then

$$\sigma = V (2\pi mkT)^{3/2}. \quad \dots (59)$$

We are less interested in the distribution of the *components* of the momenta (or of the velocities) than in that of the *magnitudes*. If the momenta are represented as radius vectors, we inquire how many end points fall in the region between  $r$  and  $r + \Delta r$ . Obviously, this number is obtained by adding all the points in the individual elements  $\Delta x \Delta y \Delta z$  included within this shell. This summation is very simple since, according to (53), the density of distribution depends only upon

$p_x^2 + p_y^2 + p_z^2 = p^2$ , and so is constant over the surface of a sphere. We thus have the sum given by the distribution function multiplied by the total number of elements contained in the spherical shell of this momentum space, i.e. simply the distribution function into the volume of the spherical shell:

$$\Delta n = \frac{N e^{-p^2/2mkT} \cdot 4\pi p^2 \Delta p}{V(2\pi mkT)^{3/2}}. \quad \dots \quad (60)$$

If we express  $p$  in terms of  $c$  and also set  $N/V = n$  we obtain the final formula

$$\frac{\Delta n}{n} = \frac{4c^2 e^{-mc^2/2kT} \Delta c}{\sqrt{\pi} \left(\frac{2kT}{m}\right)^{3/2}} = f(c) \Delta c. \quad \dots \quad (61)$$

The distribution function  $f(c)$  is plotted in fig. 1 for the case of nitrogen. Since this function vanishes at  $c = 0$  and at  $c = \infty$ , it must

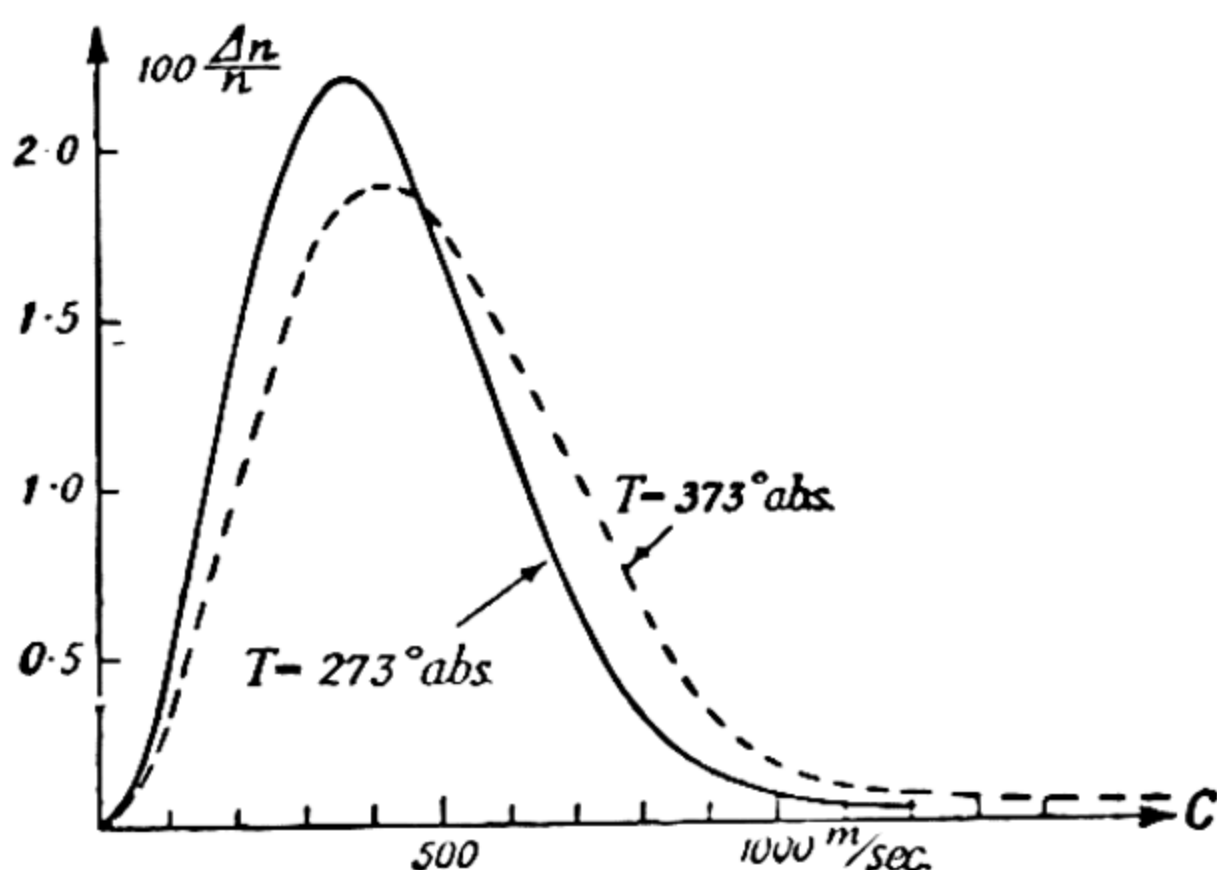


Fig. 1

have an extremum at some point. In order to locate this extremum—which, in this case, is a maximum—we introduce the notation

$$a = \sqrt{\frac{2kT}{m}} \quad \dots \quad (62)$$

for convenience. Setting the first derivative equal to zero gives

$$\frac{df}{dc} = 0 = \frac{8c}{a^3 \sqrt{\pi}} e^{-c^2/a^2} - \frac{8c^3}{a^5 \sqrt{\pi}} e^{-c^2/a^2},$$

from which

$$c_m = a = \sqrt{\frac{2kT}{m}}. \quad \dots \quad (63)$$

This may be termed the most probable molecular velocity. More-



over, we also wish to compute the mean velocity. This is given by

$$\begin{aligned} c &= \frac{1}{n} \sum c \Delta n = \frac{4}{a^3 \sqrt{\pi}} \int_0^\infty c^3 e^{-c^2/a^2} dc = \frac{4a}{\sqrt{\pi}} \int_0^\infty x^3 e^{-x^2} dx \\ &= \frac{2a}{\sqrt{\pi}} \int_0^\infty x^2 e^{-x^2} d(x^2) = -\frac{2a}{\sqrt{\pi}} \left[ x^2 e^{-x^2} \right]_0^\infty + \frac{2a}{\sqrt{\pi}} \int_0^\infty e^{-x^2} \cdot 2x dx \\ &= \frac{2a}{\sqrt{\pi}} = \frac{2}{\sqrt{\pi}} c_m = 1.128 c_m. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (64) \end{aligned}$$

Finally, the mean square velocity is

$$\begin{aligned}\overline{c^2} &= \frac{1}{n} \sum c^2 \Delta n = \frac{4}{\alpha^3 \sqrt{\pi}} \int_0^\infty c^4 e^{-c^2/\alpha^2} dc = \frac{2\alpha^2}{\sqrt{\pi}} \int_0^\infty x^3 e^{-x^2} d(x^2) \\&= -\frac{2\alpha^2}{\sqrt{\pi}} [x^3 e^{-x^2}]_0^\infty + \frac{3\alpha^2}{\sqrt{\pi}} \int_0^\infty x e^{-x^2} d(x^2) = -\frac{3\alpha^2}{\sqrt{\pi}} [x e^{-x^2}]_0^\infty \\&\quad + \frac{3\alpha^2}{\sqrt{\pi}} \int_0^\infty e^{-x^2} dx = \frac{3\alpha^2}{2\sqrt{\pi}} \int_{-\infty}^{+\infty} e^{-x^2} dx = \frac{3\alpha^2 \sqrt{\pi}}{2\sqrt{\pi}} = \frac{3\alpha^2}{2} \\&= \frac{3}{2} c_m^2; \quad \sqrt{\overline{c^2}} = 1.225 c_m.\end{aligned}$$

We have thus derived the relation between average velocity and root mean square velocity which was given without proof in Chap. XXXIV.

**(d) The equation of state of an ideal gas; Boltzmann Constant and Gas Constant**

If the value obtained above for the function  $\sigma$  of an ideal gas is inserted in the equation for the free energy (p. 588), one obtains

$$F = -kNT \log \frac{\sigma}{[h]^3} = -kNT \log \{V (2\pi mkT)^{3/2}\} + kNT \log [h]^3. \quad (66)$$

Using the second equation of (34) (p. 527), viz.

$$p = - \left( \frac{\partial F}{\partial V} \right)_T, \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad (67)$$

we obtain

$$p = \frac{kNT}{V}, \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad (68)$$

or for one mol ( $N = L$ ),

$$pv = kLT = RT. \quad . \quad . \quad . \quad . \quad . \quad (69)$$

Therefore, as already mentioned,  $k$  is the gas constant per molecule:

$$k = \frac{R}{L} = \frac{8.313 \times 10^7 \text{ erg/deg/mol}}{6.02 \times 10^{23}} = 1.380 \times 10^{-16} \text{ erg/deg.} \quad (70)$$

(e) *Osmotic pressure*

On p. 547 this phenomenon was treated from a thermodynamic point of view. It is possible to attain the same result by a statistical approach.

Unlike the pressure exerted by a gas, evident as the result of molecular impacts with the walls, osmotic pressure is less readily visualized. We can think of the membrane as a field of force affecting only the molecules of the solute—an assumption not in the least artificial since, in the last analysis, the effect of the walls of a vessel of gas is determined by the field of force of the molecules constituting the walls. This assumption is particularly apt in the case of a membrane, for there must be a certain thickness of membrane material within which the selective permeability comes about. A particle approaching the wall will then obtain a potential energy  $u$  which rises sharply toward infinity as penetration occurs. The force exerted on the particle by the wall is given by

$$\mathbf{F} = -\text{grad } u. \quad . . . . . (71)$$

Inasmuch as  $u$  increases as the particle approaches the wall,  $\mathbf{F}$  must be directed into the solution, and according to Newton's Third Law the particle exerts an equal and opposite force on the wall. Qualitatively at least, the situation is the same as for the phenomenon of gas pressure. It remains to show that the gas laws actually hold for osmotic pressure.

For the present, assume that equilibrium has not yet been reached on the solution side. In this condition the full pressure excess does not yet exist on the solution side, and the force field of the membrane repels the particles. Now, it is known that when a molecule or an ion moves in a liquid, viscosity causes some of the solvent to be dragged along (see p. 419). The liquid thus streaming away from the membrane is replaced by fresh solvent which passes through the membrane, for no barrier exists in so far as the solvent is concerned. The solvent streams in until the pressure excess equals the osmotic pressure. At equilibrium the Boltzmann distribution law holds, which means that the number of particles of energy  $u$  in each cubic centimetre is given by

$$n = n_0 e^{-u/kT}. \quad . . . . . (72)$$

Here  $n_0$  is the number of particles per cubic centimetre at a point in the solution far from the membrane. The force on an element of volume of the solution containing  $n$  particles of energy  $u$  per cubic centimetre is

$$f dv = -n \text{ grad } u dv. \quad . . . . . (73)$$

It follows from (72) that

$$\text{grad } n = -\frac{n_0}{kT} e^{-u/kT} \text{grad } u = -\frac{n}{kT} \text{grad } u. \quad (72')$$

Substituting into (73), the resulting expression for the force per unit volume becomes

$$f = kT \text{grad } n. \quad (74)$$

According to the fundamental hydrostatic equation (p. 191), we have, at equilibrium,

$$\text{grad } p = f = kT \text{grad } n. \quad (75)$$

Integrating,

$$p = kTn + p_0, \quad (76)$$

where  $p_0$  is the pressure corresponding to  $n = 0$ , i.e. the pressure on the solvent side. The osmotic pressure is  $p - p_0$ . If one mol of solute is dissolved in a volume  $v$  of solution, then  $n = L/v$  and

$$p_{\text{osm}} = \frac{kL}{v} T = \frac{RT}{v}. \quad (77)$$

Thus we have deduced the van't Hoff law (p. 547) on a statistical basis.

## 6. The Law of Equipartition and its Application to the Specific Heats.

The energy of a single particle is, in most cases occurring in practice, a homogeneous function of the squares of a number of position or momentum co-ordinates to which we give the unified designation  $x_1, x_2 \dots x_f$ . The total number  $f$  of parameters upon which the energy depends is called the number of degrees of freedom. This is unlike the usage in mechanics, where the number of position co-ordinates alone was called the number of degrees of freedom of the system. Under the condition that we have a homogeneous quadratic energy function, Euler's theorem (p. 531) is valid:

$$x_1 \frac{\partial u}{\partial x_1} + x_2 \frac{\partial u}{\partial x_2} + \dots + x_f \frac{\partial u}{\partial x_f} = 2u. \quad (78)$$

Passing directly to integration instead of summation, the total energy of the system is given by

$$U = \int u dN = \frac{N \int \int \int \dots e^{-u/kT} u dx_1 dx_2 dx_3 \dots dx_f}{\int \int \int \dots e^{-u/kT} dx_1 dx_2 dx_3 \dots dx_f} = \frac{NZ}{\sigma}. \quad (79)$$



The numerator  $Z$  may be transformed with the help of Euler's theorem. We have

$$\begin{aligned} 2Z = & \int \int \int \int \dots e^{-u/kT} x_1 \frac{\partial u}{\partial x_1} dx_1 dx_2 \dots dx_f \\ & + \int \int \int \int \dots e^{-u/kT} x_2 \frac{\partial u}{\partial x_2} dx_1 dx_2 \dots dx_f \\ & + \dots + \int \int \int \int \dots e^{-u/kT} x_f \frac{\partial u}{\partial x_f} dx_1 dx_2 \dots dx_f. \end{aligned} \quad (80)$$

In the first integral we integrate with respect to  $x_1$  by parts:

$$\int e^{-u/kT} x_1 \frac{\partial u}{\partial x_1} dx_1 = -kT [x_1 e^{-u/kT}] + kT \int e^{-u/kT} dx_1. \quad (81)$$

In general, the limits are  $-\infty$  and  $+\infty$ —at any rate, there are two limits of equal magnitude in  $x_1$ , but of opposite sign. Then, since  $u$  is proportional to  $x_1^2$ , the bracketed expression vanishes when the limits are substituted, so that only the integral remains on the right side. Substituting this in (79), the first term yields  $kT\sigma$ . If we proceed in this way with every term, we obtain

$$2Z = f k T \sigma. \quad \dots \dots \dots (82)$$

Hence

$$U = \frac{N f k T}{2} = N f \bar{u}. \quad \dots \dots \dots (83)$$

This is the celebrated Law of the Equipartition of Energy in classical statistical mechanics: *Each degree of freedom which enters quadratically into the expression for the energy of the entire system contributes, on the average,  $kT/2$  to this energy.*

This theorem may be verified in a simple manner by applying it to the calculation of the specific heats of gases. Since

$$c_v = \left( \frac{\partial u}{\partial T} \right)_v, \quad \dots \dots \dots (84)$$

the molecular heat becomes, according to (83),

$$c_v = \frac{L k f}{2} = \frac{f R}{2}. \quad \dots \dots \dots (85)$$

Since  $R$  is approximately equal to 2 calories per degree,

$$c_v \approx f.$$

The degrees of freedom of a monatomic gas are the three components

of the translational velocity; the specific heat per mol is then

$$c_v = \frac{3R}{2}, \quad . . . . . (86)$$

whence 
$$c_p = c_v + R = \frac{5R}{2}, \quad . . . . . (87)$$

and 
$$\frac{c_p}{c_v} = \frac{5}{3} \quad . . . . . (88)$$

for any monatomic gas. These relations have been verified from the lowest to the highest attainable temperatures.

As a model of a diatomic molecule we take a rigid dumb-bell.\* In addition to translational energy, the molecule has rotational energy. The rotation, too, may be resolved into components about three axes. The rotation about the figure axis itself does not contribute to the kinetic energy, since the moment of inertia about this axis is zero; hence there are two additional degrees of freedom of rotation for a diatomic gas, and we have

$$c_v = \frac{5R}{2}, \quad \frac{c_p}{c_v} = \frac{7}{5} \quad . . . . . (89)$$

This relation is also generally valid at room temperature. Nevertheless, it has been observed—especially for hydrogen—that the specific heat drops to the value for monatomic gases as the temperature is lowered. This does not agree with the Law of Equipartition.

Molecules consisting of more than two atoms have three degrees of rotational freedom, so that as long as we consider such molecules rigid (internal vibrations absent), the specific heat at constant volume is given by

$$c_v = \frac{6R}{2}, \quad \frac{c_p}{c_v} = \frac{4}{3}, \quad . . . . . (90)$$

which holds for any such polyatomic molecule.

As we saw on p. 575, the atoms of a solid body are bound to their equilibrium positions by quasi-elastic restoring forces. Thus an atom possesses not only kinetic energy but also potential energy which, by the results of p. 577, is also a quadratic function of the position co-ordinates. This means that for an atom of a solid there are six

\* In reality the two atoms may also vibrate in the line joining them. The reason that these vibrations contribute nothing to the specific heat at ordinary temperatures will be seen in the next chapter.

degrees of freedom contributing to the energy, whence

$$c_v = \frac{6R}{2} \approx 6. \quad . \quad . \quad . \quad . \quad . \quad . \quad (91)$$

This is the same as the empirical Dulong-Petit Law of Atomic Heats. At room temperature this law is in general well obeyed—only the lighter elements (e.g. diamond) show departures. At lower temperatures, however, the specific heats decrease—as a matter of fact, to zero—which is what the Nernst Heat Theorem requires. But this decrease is also contrary to the Law of Equipartition. Thus it seems as if this law is valid at higher temperatures, but that at lower temperatures an ankylosis\* of degrees of freedom takes place. This indicates that the classical statistics must be modified in some way.

### 9. Fluctuations. Transition from Microscopic to Macroscopic Motion.

If thermodynamical equilibrium, corresponding to the maximum value of the entropy, is of a statistical nature only, departures from the most probable values are to be expected when very small regions are observed. For example, if it were possible to observe an element of volume containing but one hundred molecules we would observe relatively large fluctuations of density in this space; if but one additional molecule should enter, the density would be increased by 1 per cent. These variations in density are responsible for the scattering of light in the atmosphere. A knowledge of the theory of this scattering makes it possible to calculate Avogadro's Number from the spectral energy distribution of skylight. Again, imagine a small, movable surface element in the wall of a vessel containing a gas. Let the element be so small that but a few hundred molecules strike it each second. While the element would not respond to each impact separately, the likelihood of considerable pressure fluctuations is no longer small, and an irregular fluttering of the element is to be expected. This phenomenon was discovered about a century ago by the English botanist Brown, but its nature remained unexplained for a long time. If we observe very small, barely visible particles suspended in a liquid (colloid particles), these exhibit an irregular vibratory motion because of the fact that the impacts of the molecules of the liquid do not exactly balance at each instant, so that the particle is urged sometimes in one direction, sometimes in another. Looking upon the colloid particle as a large molecule, its kinetic energy must, on the average, be equal to  $kT/2$ , according to the Law of Equipartition. What we observe in the microscope, however, are velocities of much lower order of magnitude, if we take the velocity to be the distance covered divided by the

\* Poincaré borrowed this very apt term from pathology to denote the "freezing up" of a degree of freedom.



time. In reality, however, the velocity alters its direction so often that the observed motion is merely a smoothed-over version of the actual zigzag path. The path of such a particle is suggestive of a type of curve occurring in higher mathematics—a curve which is everywhere continuous but yet has no tangent. It is quite hopeless to attempt—say by cinematography—to ascertain the details of the motion in order to determine the actual mean kinetic energy. Rather, if we wish to give a theory of the phenomenon which is amenable to experimental test, we must consider the mean displacement instead of the velocity. Consider only displacements in one dimension. Let there be  $N$  particles per unit volume between  $x$  and  $x + dx$  at the time  $t = 0$ . After a given time  $\tau$  has elapsed, we consider a volume element of the same size at the point  $x'$ . Particles not within this volume element at first, but now inside it, have entered from the neighbouring elements. The probability that a particle enter from an adjoining element is, of course, a function of the distance  $x' - x$ , and the difference in time  $\tau$  of the two observations. Accordingly, we denote it by  $f_\tau(x' - x)$ . The function  $f_\tau$  includes also the case in which the particles were previously in the element  $x'$  if we set  $x' - x = 0$ . Since the particles must come from some volume element, the density at  $x'$  at the time  $\tau$  is

$$N(x', \tau) = \int_{-\infty}^{+\infty} N(x, 0) f_\tau(x' - x) dx. \quad \dots (92)$$

Introducing the displacement  $X = x - x'$ ,

$$N(x', \tau) = \int_{-\infty}^{+\infty} [N(x' + X, 0)] f_\tau(X) dX. \quad \dots (93)$$

The function  $f_\tau$  is even, since positive and negative displacements are equally probable:

$$f_\tau(X) = f_\tau(-X). \quad \dots (94)$$

We now develop the left member in powers of  $\tau$ , the right in powers of  $X$ :

$$\begin{aligned} N(x', 0) + \tau \frac{\partial N}{\partial t} + \dots \\ = \int_{-\infty}^{+\infty} \left\{ N(x', 0) + X \frac{\partial N}{\partial x} + \frac{1}{2} X^2 \frac{\partial^2 N}{\partial x^2} + \dots \right\} f_\tau(X) dX \dots (95) \end{aligned}$$

The derivatives  $\partial N / \partial t$ ,  $\partial N / \partial x \dots$  have fixed values for the time  $t = 0$  and the point  $x'$ . On account of the symmetry of the function  $f_\tau(X)$ , the integral

$$\int_{-\infty}^{+\infty} X f_\tau(X) dX$$

vanishes; further, since each particle which enters must come from some other element, the integral

$$\int_{-\infty}^{+\infty} f_{\tau}(X) dX = 1.$$

Dividing by  $\tau$ , we obtain from equation (95):

$$\frac{\partial N}{\partial t} = \frac{1}{2\tau} \frac{\partial^2 N}{\partial x^2} \int_{-\infty}^{+\infty} X^2 f_{\tau}(X) dX = \frac{1}{2\tau} \frac{\partial^2 N}{\partial x^2} \overline{X^2}. \quad (96)$$

The integral on the right represents the mean square displacement  $\overline{X^2}$ , since it is the sum of the squares of the displacements, each being multiplied by the probability of its occurrence. Now it is readily shown that the mean square of the displacement is proportional to the time interval of observation, so that  $\overline{X^2}/2\tau$  is constant: if we observe the instantaneous positions of a particle at equal time intervals  $\tau$  which are not too short, we may look upon the individual displacements as independent, on account of the random nature of the impacts.\* After a series of  $p$  observations the resulting displacement is

$$X^{(p)} = X_1 + X_2 + X_3 + \dots + X_p. \quad (97)$$

The square of the displacement is then

$$(X^{(p)})^2 = (X_1 + X_2 + \dots + X_p)^2 = \sum_1^p X_r^2 + 2 \sum_1^p \sum_1^p X_r X_s. \quad (98)$$

If we now average over the displacements experienced by a large number of particles, the double sum will vanish on account of the equal probability of positive and negative displacements, and so we have

$$\overline{(X^{(p)})^2} = p \overline{X^2}. \quad (99)$$

This means that after the time  $p\tau$  has elapsed we observe a mean square displacement of magnitude  $p\overline{X^2}$ , i.e.  $p$  times the value corresponding to the time  $\tau$ . Thus  $\overline{X^2}$  is proportional to the interval of observation.

Equation (96) is neither more nor less than the well-known differential equation of diffusion, which is analogous to the equation of heat conduction, concentration taking the place of temperature. Thus, if the concentration varies from place to place, a stream of particles will flow from the points of high concentration to those of

\* If the observations are made too frequently, a given velocity of the particle will still persist after the instant of observation and so will affect the following observation. In a practical case, however, it is hardly possible to make observations at such short intervals.

low concentration. We set the *density* of this flow proportional to the slope of the concentration, i.e. to  $-D \text{ grad } N$ , or in the one-dimensional case,  $-D(\partial N/\partial x)$ . The change in the number of particles within a volume element  $dx dy$  is due to the entry of particles from one side and the departure of particles from the other. At the point  $x$  a current

$$\left(-D \frac{\partial N}{\partial x}\right)_x dy$$

enters and at the point  $x + dx$  the current

$$\left(-D \frac{\partial N}{\partial x}\right)_{x+dx} dy$$

leaves. The excess of the number entering over the number leaving is the rate at which the particles are diffusing into the volume element, so that

$$\frac{\partial N}{\partial t} dx dy = D \frac{\partial^2 N}{\partial x^2} dx dy \quad \text{or} \quad \frac{\partial N}{\partial t} = D \frac{\partial^2 N}{\partial x^2}. \quad (100)$$

Thus (cf. (96))

$$D = \frac{\overline{X^2}}{2\tau} \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad (101)$$

represents the diffusion coefficient of the particles. If the particles are in a field of force, e.g. the earth's gravitational field, the Boltzmann equilibrium distribution sets in. This is given by

$$N = N_0 e^{-u/kT} \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad (102)$$

If the force is constant, the potential energy is given by  $Fx$ . We may imagine the equilibrium to result from the motion of the particles due to the action of the force together with a diffusion current which seeks to satisfy (102). The velocity of a particle which is in equilibrium under the action of the applied force and viscosity is given by Stokes's Law (p. 222):

$$v = \frac{F}{6\pi\eta a}, \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad (103)$$

where  $a$  is the radius of the particle and  $\eta$  is the coefficient of viscosity of the solvent. The density of this stream of particles is then

$$s = \frac{NF}{6\pi\eta a}, \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad (104)$$

But a diffusion current of equal strength flows in the opposite direction, tending to keep the distribution of concentration constant. The



density of this flow is given by  $-D(\partial N/\partial x)$ . Equating these two flow densities,

$$-D \frac{\partial N}{\partial x} = \frac{NF}{6\pi\eta a} \quad \cdot \cdot \cdot \cdot \cdot \cdot (105)$$

On the other hand, it follows from (102), for  $u = Fx$ , that

$$\frac{1}{N} \frac{\partial N}{\partial x} = -\frac{F}{kT} \quad \cdot \cdot \cdot \cdot \cdot \cdot (106)$$

Inserting this expression in (105) and using (101), we obtain the important equation

$$\frac{D}{kT} = \frac{1}{6\pi\eta a}, \quad \cdot \cdot \cdot \cdot \cdot \cdot (107)$$

or

$$\frac{\overline{X^2}}{\tau} = \frac{kT}{3\pi\eta a} \quad \cdot \cdot \cdot \cdot \cdot \cdot (108)$$

This is a general expression for the mean square of the displacement of a particle.\*

\* The above theory was developed by Einstein, and the final formula (108) was experimentally checked by Perrin and others. Consult A. Einstein, *Ann. der Physik*, **17**, p. 549 (1905); E. Bloch, *The Kinetic Theory of Gases*, English translation by P. A. Smith (London, Methuen & Co., 1924). Einstein also developed a formula for the rotational Brownian movement experienced by particles of irregular shape. This formula, too, was verified experimentally. This is discussed by Bloch, *loc. cit.*

## CHAPTER XXXVI

### THE CLASSICAL QUANTUM STATISTICS

#### 1. The Quantum Condition for the Subdivision of the Phase Space of an Oscillator.

As we saw in § 6 (p. 595) of the preceding chapter, the Law of Equipartition leads to conflict with experiment when applied to the calculation of specific heats. The results are entirely incorrect when the theorem is applied to a volume filled with radiation (see p. 622). In connexion with this problem, Planck was led to advance the hypothesis, in 1900, that a simple, linear, harmonic oscillator can absorb energy only in integral multiples of a minimum quantity of energy which is proportional to the frequency.\* Thus the energy possessed by such a system is always an integral multiple of the quantity  $h\nu$ :

$$u_n = nh\nu. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

The factor of proportionality,  $h$ , is the most important of all constants in atomic physics. It occurs universally, and is not connected merely with the linear oscillator, as might be inferred from the manner in which it was introduced. Its numerical value is

$$h = 6.62 \times 10^{-27} \text{ erg sec.}$$

This constant is usually referred to as the Planck quantum of action, since it has the dimensions of action (energy  $\times$  time). We encountered this constant for the first time in the theory of discharges in gases (p. 432). The greater the natural frequency  $\nu$ , the greater are the intervals in energy content, and the greater is the difference between the quantum theory and the classical statistics, where a continuous series of energy states was tacitly assumed.

What is the significance of the Planck hypothesis for the phase space of a system consisting of oscillators? It is obvious that for an arbitrary division into cells there can no longer be equal probability for cells of equal size, for if we mark the points in which the oscillator has exactly the energy  $nh\nu$ , it will always be possible to find volume elements containing no points (i.e. elements whose probability is zero) if the subdivision is arbitrary. Nevertheless, it is possible to preserve

\* M. Planck, *Ann. der Physik*, 4, p. 553 (1901).

the validity of the Liouville theorem by subdividing the phase space in a particular way. For the linear oscillator, which has but one position co-ordinate and one momentum component, the phase space is the  $pq$ -plane. The potential and kinetic energies are

$$\frac{1}{2}fq^2 \quad \text{and} \quad \frac{1}{2m}p^2$$

respectively. In an energy state in which the total energy is  $n h \nu$  we have the relation

$$n h \nu = \frac{1}{2}fq^2 + \frac{1}{2m}p^2, \quad . \quad . \quad . \quad . \quad . \quad (2)$$

or

$$\frac{q^2}{\frac{2n h \nu}{f}} + \frac{p^2}{2m n h \nu} = 1. \quad . \quad . \quad . \quad . \quad . \quad (2')$$

This is the equation of an ellipse whose semi axes are

$$\sqrt{\frac{2n h \nu}{f}} \quad \text{and} \quad \sqrt{2m n h \nu}.$$

This ellipse is termed the *phase orbit*, since the phase point traverses this path once during a complete vibration of the oscillator, as long as the energy remains constant. The area enclosed by the phase orbit is given by the phase integral (cf. p. 129):

$$J = \oint p dq = ab\pi = 2\pi n h \nu \sqrt{\frac{m}{f}}. \quad . \quad . \quad . \quad . \quad (3)$$

Since

$$\nu = \frac{1}{2\pi} \sqrt{\frac{f}{m}},$$

this becomes

$$J = n h. \quad . \quad . \quad . \quad . \quad . \quad (4)$$

*In the case of the oscillator, the Planck hypothesis leads to the condition that only those states occur for which the value of the phase integral is an integral multiple of  $h$ . This is the general formulation of the quantum condition. It is only by chance that for the oscillator this condition is equivalent to the energy condition; for other systems the two conditions are quite distinct. If we now draw the ellipses corresponding to  $n = 1, 2, 3 \dots$  we obtain a division of the plane into elliptical rings, each of area  $h$ . If we ascribe to each of these rings the energy value corresponding to its inner boundary, the innermost ring has the energy 0, the next one  $1 h \nu$ , &c. Thus we again arrive at a division of the phase space into elements of equal size; if all quantum states are looked upon as equally likely, each element has the same*



a priori probability and we can proceed as in the classical statistics.

If the elementary particle has several co-ordinates, the condition that the phase integrals be integers holds for each co-ordinate, and the element of phase space becomes  $h^s$ . Thus the value of the dimensional constant  $[h]$  introduced on p. 588 is determined. We write  $h$  for it in the future.

The chief difference between the present procedure and the classical statistics is that the prescribed division of phase space into cells of definite size no longer permits the replacement of the partition sum  $\sigma$  by an integral, since the latter implies a continuous change of energy. We thus obtain for the partition sum of the oscillator (p. 587)

$$\sigma = h \sum e^{-nh\nu/kT} = h(1 + e^{-h\nu/kT} + e^{-2h\nu/kT} + e^{-3h\nu/kT} + \dots). \quad (5)$$

The sum of this geometrical progression is

$$\sigma = \frac{h}{1 - e^{-h\nu/kT}} \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad (6)$$

This expression for  $\sigma$  has important applications to specific heats, as will be shown in the following section.

## 2. Temperature Variation of the Vibrational Contribution to the Specific Heats.

### (a) Diatomic molecules

The two atoms which constitute a diatomic molecule are not rigidly connected to one another, but execute vibrations about their common centre of gravity in the direction of the axis of figure. The evidence for these vibrations is obtained from observations on band spectra (cf. p. 680). Why do not the two degrees of freedom of vibrational kinetic and potential energy contribute to the specific heat, which apparently comes out correctly when only the translational and rotational degrees of freedom are taken into account? The quantum theory furnishes information on this point. With regard to the vibrational contribution to the total energy, we can consider the diatomic gas to be a system of linear oscillators, since only vibrations along the axis of figure of each molecule are possible. We have already calculated the partition sum for an oscillator. The free energy of one mol of such oscillators is then (p. 588)

$$f = -kLT \log \frac{\sigma}{h} = -RT \log h + RT \log(1 - e^{-h\nu/kT}) + RT \log h, \quad (7)$$

whence, by differentiation,

$$\left(\frac{\partial f}{\partial T}\right)_v = R \log(1 - e^{-h\nu/kT}) - \frac{RTe^{-h\nu/kT} h\nu}{(1 - e^{-h\nu/kT}) kT^2} \quad \cdot \quad (8)$$

The energy then becomes

$$u = f - T \left( \frac{\partial f}{\partial T} \right)_v = \frac{Rh\nu e^{-h\nu/kT}}{k(1 - e^{-h\nu/kT})} = \frac{Rh\nu}{k} \frac{1}{(e^{h\nu/kT} - 1)}. \quad (9)$$

For high temperatures, the exponential function in the denominator may be developed in a power series ending with the second term, thus giving

$$u = RT = L \cdot 2 \frac{kT}{2}. \quad \cdot \cdot \cdot \cdot \cdot \quad (10)$$

But this is the value of the energy corresponding to the two degrees of freedom according to the equipartition of energy. *The classical statistics remains valid when the temperature is high*; at low temperatures the finite differences of the quantum states manifest themselves by the fact that their contributions are not as large as the classical value. The contributions vanish entirely for very small values of  $T$ , and this degree of freedom is "ankylosed" (p. 596). It is evidently a question of the ratio of the temperature to a temperature  $\Theta$  which is characteristic of the oscillator:

$$\Theta = \frac{h\nu}{k}. \quad \cdot \cdot \cdot \cdot \cdot \quad (11)$$

We have

$$\begin{aligned} c_v &= \left( \frac{\partial u}{\partial T} \right)_v = R\Theta \frac{d}{dT} \frac{1}{e^{\Theta/T} - 1} = R \left( \frac{\Theta}{T} \right)^2 \cdot \frac{e^{\Theta/T}}{(e^{\Theta/T} - 1)^2} \\ &= R\Phi \left( \frac{\Theta}{T} \right), \quad \cdot \cdot \cdot \cdot \cdot \quad (12) \end{aligned}$$

where we introduce the function

$$\Phi(x) = \frac{x^2 e^x}{(e^x - 1)^2}. \quad \cdot \cdot \cdot \cdot \cdot \quad (12a)$$

For small values of  $x$ , i.e. for high temperatures, the value of  $\Phi$  approaches unity, while for large values of  $x$ , i.e. for low temperatures,  $\Phi$  vanishes. At room temperature (about  $300^\circ$  abs) the state of affairs is this:

$$\begin{aligned} \text{For } x > 10, \Phi(x) < 0.005; \Theta > 3000^\circ, \\ \text{i.e. } \nu > 6.2 \times 10^{13} \text{ sec.}^{-1} \text{ or } \lambda < 4.8 \mu. \end{aligned}$$

$$\begin{aligned} \text{For } x < 0.5, \Phi(x) > 0.975; \Theta < 150^\circ, \\ \text{i.e. } \nu < 3.1 \times 10^{12} \text{ sec.}^{-1} \text{ or } \lambda > 95 \mu. \end{aligned}$$

The accompanying wave-lengths are those of electromagnetic waves of the given frequencies; these are measured spectroscopically as

the *infra-red characteristic radiation* of the system. Thus we see that an infra-red characteristic vibration whose wave-length is less than  $4.8 \mu$  is practically absent at room temperature. This is actually the case for diatomic gases. The function  $\Phi(x)$  steadily decreases from 1 to 0 as  $x$  increases from 0 to  $\infty$ . Hence, according to equation (12) there must be an increase of the specific heat as the temperature is raised. This is observed for HCl, in good agreement with the theory. If we insert in (12) the frequency corresponding to the infra-red characteristic vibration at  $\lambda = 3.47 \mu$ , there results

$T$	$c_v$ calc.	$c_v$ obs.
1500°	5.41	5.45
2000°	5.66	5.68
2500°	5.82	5.90

(b) *The solid state*

If we represent the structure of a solid by atoms vibrating about centres of equilibrium, we must remember that there are three directions of vibration; we used this fact in deriving the Law of Dulong and Petit. This means that one mol of a solid substance is equivalent to  $3L$  oscillators, and we obtain for the molecular heat of a solid, exactly as above,

$$c_v = 3R \left( \frac{\Theta}{T} \right)^2 \frac{e^{\Theta/T}}{(e^{\Theta/T} - 1)^2} = 3R \Phi \left( \frac{\Theta}{T} \right). \quad . \quad . \quad (13)$$

In this instance the characteristic temperature  $\Theta$  is generally so low that all vibrations are fully stimulated at room temperature, and so the limiting value  $3R$  holds for  $c_v$ . Since the mass occurs in the denominator of the formula for  $\nu$ , it is only for the lightest elements that the characteristic frequency is so great that the full value of  $c_v$  is not attained at room temperature.

These simple considerations give a good qualitative representation of the variation of the specific heat of a solid, but cannot be connected with an individual characteristic vibration. If we take several characteristic frequencies, we obtain so many free constants that the formula is merely of the nature of a good interpolation formula. It is evident that the assumption of a single frequency resulting from independent oscillators cannot lead to quantitatively correct results, for the oscillators are by no means independent. In fact, a disturbance applied to one lattice point is transmitted to the neighbouring points, for we are dealing with a system of coupled oscillators. The calculation for



this case is rather troublesome, but has been carried out by Born and Thirring for simple lattices.\*

*Ex. 129.* Determine, according to the conditions given above, the quantum levels of the rotational motion of a rigid dumb-bell (diatomic molecule) of moment of inertia  $I$ . Investigate the falling off in the rotational part of the specific heat of  $H_2$ .

### 3. The Debye Theory of the Specific Heats of Solids.

The law of variation of the specific heats of solids is obtained with much less calculation by a method due to Debye. In Debye's theory the atomistic structure is thrust into the background and the coupled vibrations of the system of oscillators are considered to be elastic waves which traverse the body and are reflected at its boundaries, thus giving a system of standing waves. A wave corresponding to a given frequency has the same determining properties as a linear oscillator; firstly, a frequency  $\nu$ , and secondly, two energy components, corresponding to the potential and kinetic energy of the oscillator. For elastic waves there is a part corresponding to compression and one to motion; for electromagnetic waves we have the energy densities of the electric and magnetic fields. These properties completely determine a longitudinal wave, but for transverse waves the plane of polarization of the vibrations must be specified also. This may be determined by resolving the amplitude along two mutually perpendicular directions. Thus an unpolarized transverse wave corresponds to two oscillators, one for each component of the vibration.

Following a method given by Rayleigh and Jeans, we now proceed to calculate the number of standing waves which can be contained in a body whose volume is  $V$ . For simplicity, let this be a cube of edge  $l$ .

Imagine a wave travelling parallel to the  $x$ -axis. This wave will travel to and fro between the walls which are parallel to the  $yz$ -plane. A system of standing waves results from the superposition of the incident and reflected waves. There is either a node or a loop (antinode) at the wall, depending upon the nature of the waves and of the walls. It is immaterial, in the following computation, whether one or the other condition exists. For electromagnetic waves and metallic reflecting walls, for example, there must be a node of the vector  $\mathbf{E}$  at the wall. In the problem of the specific heats we deal with elastic waves representing the vibrations of the lattice points. Only those waves can travel in the  $x$ -direction for which  $l$  is an integral multiple of half the wave-length (cf. fig. 1). Calling the phase velocity  $c$ , we have

$$l = \frac{n\lambda}{2} = \frac{nc}{2\nu} \quad \cdot \cdot \cdot \cdot \cdot \cdot \quad (14)$$

\* Consult M. Born, *The Mechanics of the Atom*, Eng. trans and revision by J. Fisher and D. Hartree (London, G. Bell & Sons, 1927).

For a cubical body, the same relationships are valid for waves in the  $y$ - and  $z$ -directions. But these are not the only possible standing waves. Consider, for example, a wave whose normal

$\mathbf{n} = +j \cos \beta + k \cos \gamma$  lies in the  $yz$ -plane. This wave will be reflected from the  $xy$ -plane, and we must consider the reflected wave as well as the incident wave. Now the superposition of the two is, by p. 64, equivalent to a wave moving parallel to the  $y$ -axis with the phase velocity  $c' = c/\cos \beta$ , whose amplitude in the direction of the  $z$ -axis is modulated with the wave-length  $\lambda_z' = \lambda/\cos \gamma$ . The boundary condition for the planes parallel to the  $xz$ -plane thus demands that

$$\frac{2l\nu}{c'} = \frac{2l\nu \cos \beta}{c}$$

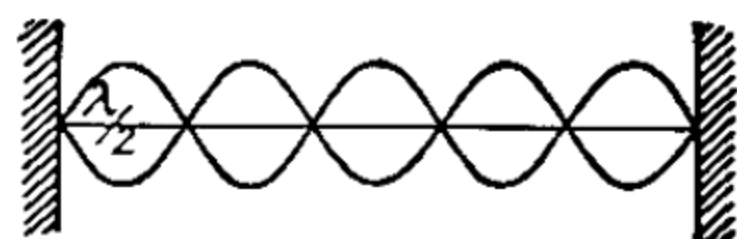


Fig. 1

should be a whole number, and the boundary condition for the faces parallel to the  $xy$ -plane requires similarly that

$$\frac{2l}{\lambda_z'} = \frac{2l\nu \cos \gamma}{c}$$

should be an integer.

Consider now a wave whose normal direction is quite arbitrary:  $\mathbf{n} = \cos \alpha \mathbf{i} + \cos \beta \mathbf{j} + \cos \gamma \mathbf{k}$ . The superposition of the incident wave and that reflected from the  $yz$ -plane yields

$$\begin{aligned} A + A' &= A_0 e^{i\omega(t - (x \cos \alpha + y \cos \beta + z \cos \gamma)/c)} + A_0 e^{i\omega(t + x \cos \alpha/c - (y \cos \beta + z \cos \gamma)/c)} \\ &= 2A_0 \cos \left( \frac{\omega x \cos \alpha}{c} \right) e^{i\omega(t - (y \cos \beta + z \cos \gamma)/c)}. \end{aligned}$$

This wave, whose front is modulated in the direction of the  $x$ -axis, is now reflected at the  $xz$ -plane; superposition of the two waves then gives

$$\begin{aligned} A + A' + A'' &= 2A_0 \cos \left( \frac{\omega x \cos \alpha}{c} \right) e^{i\omega(t - y \cos \beta/c - z \cos \gamma/c)} \\ &\quad + 2A_0 \cos \left( \frac{\omega x \cos \alpha}{c} \right) e^{i\omega(t + y \cos \beta/c - z \cos \gamma/c)} \\ &= 4A_0 \cos \left( \frac{\omega x \cos \alpha}{c} \right) \cos \left( \frac{\omega y \cos \beta}{c} \right) e^{i\omega(t - z \cos \gamma/c)}. \end{aligned}$$

This wave front with lattice-like modulation, when combined with the wave reflected from the  $xy$ -plane, gives the system of standing waves

$$\begin{aligned} A + A' + A'' + A''' &= 8A_0 \cos \left( \frac{\omega x \cos \alpha}{c} \right) \cos \left( \frac{\omega y \cos \beta}{c} \right) \cos \left( \frac{\omega z \cos \gamma}{c} \right) e^{i\omega t}. \quad (15) \end{aligned}$$

The fulfilment of the boundary conditions at the walls requires that

$$\frac{2\nu l \cos \alpha}{c} = n_1, \quad \frac{2\nu l \cos \beta}{c} = n_2, \quad \frac{2\nu l \cos \gamma}{c} = n_3, \quad . \quad (16)$$

or 
$$\cos \alpha = \frac{cn_1}{2\nu l}, \quad \cos \beta = \frac{cn_2}{2\nu l}, \quad \cos \gamma = \frac{cn_3}{2\nu l}. \quad . \quad (16')$$

Squaring and adding,

$$n_1^2 + n_2^2 + n_3^2 = \frac{4\nu^2 l^2}{c^2}. \quad . \quad . \quad . \quad (17)$$

To every frequency  $\nu$  there correspond triads of positive integers which satisfy (17), and to each triad there corresponds one direction which may serve as a wave normal. We now seek to determine the number of waves between the frequency limits  $\nu$  and  $\nu + d\nu$ . In order to compute this number we need only look upon  $n_1$ ,  $n_2$ , and  $n_3$  as rectangular co-ordinates. Then (17) represents the equation of a sphere of radius  $2\nu l/c$ . Since *one* integral lattice point corresponds to each cube of edge unity, the number of allowed vibrations in the range  $\nu$  to  $\nu + d\nu$  is evidently equal to the volume of one *octant*\* of the spherical shell included between the radii  $2\nu l/c$  and  $2(\nu + d\nu)l/c$ . This is

$$dZ = \frac{1}{8} \cdot \frac{4\pi \cdot 4\nu^2 l^3 \cdot 2d\nu}{c^3} = \frac{4\pi \nu^2 l^3}{c^3} d\nu, \quad . \quad . \quad . \quad (18)$$

or, per cubic centimetre of the body, there will be

$$\frac{dZ}{V} = \frac{4\pi \nu^2}{c^3} d\nu \quad . \quad . \quad . \quad . \quad (18')$$

vibrations between the limits  $\nu$  and  $\nu + d\nu$ , and this number is independent of the particular form of the body, which fact may be rigorously demonstrated.

For the elastic waves corresponding to the thermal agitation of the atoms of a solid we have (p. 181) longitudinal as well as transverse waves. These have different velocities, which we denote by  $c_l$  and  $c_t$  respectively. On account of their two azimuths of polarization, the transverse waves must be counted twice. Accordingly, we have for the total number of waves in the volume  $v$  of one mol, between the frequencies  $\nu$  and  $\nu + d\nu$ ,

$$dz = 4\pi v \nu^2 \left( \frac{1}{c_l^3} + \frac{2}{c_t^3} \right) d\nu. \quad . \quad . \quad . \quad (19)$$

According to this formula, the number of waves increases as the square of  $\nu$  and without limit, so that the number of oscillators becomes infinite, while in reality one mol of a solid contains only  $3L$  oscillators. Debye arbitrarily assumed that the total number of waves should be

\* Since only positive integral  $n_i$  are allowed.



precisely  $3L$ . This assumption, which at first sight seems rather forced, is justified by its results. Moreover, it results quite automatically by the rigorous computation for a system of coupled oscillators (see end of the preceding section). The total number of waves is obtained by integrating over all frequency intervals from 0 to the limiting frequency  $\nu_0$ . Thus  $\nu_0$  is determined by the equation

$$\int_0^{\nu_0} 4\pi\nu^2 \left( \frac{1}{c_l^3} + \frac{2}{c_t^3} \right) d\nu = \frac{4\pi\nu}{3} \left( \frac{1}{c_l^3} + \frac{2}{c_t^3} \right) \nu_0^3 = 3L. \quad (20)$$

If we treat each wave as an oscillator we can proceed exactly as in § 2 (p. 605). The intrinsic energy of a system of  $N$  oscillators of frequency  $\nu$  is, by equation (9) (p. 606),\*

$$U = \frac{Nh\nu}{e^{h\nu/kT} - 1}.$$

Thus for  $dz$  oscillators we have an amount of energy

$$du = 4\pi\nu \left( \frac{1}{c_l^3} + \frac{2}{c_t^3} \right) \frac{h\nu^3 d\nu}{e^{h\nu/kT} - 1} = \frac{9L}{\nu_0^3} \frac{h\nu^3 d\nu}{e^{h\nu/kT} - 1}, \quad (21)$$

and the total energy of one mol is

$$u = \frac{9Lh}{\nu_0^3} \int_0^{\nu_0} \frac{\nu^3 d\nu}{e^{h\nu/kT} - 1}. \quad \cdot \cdot \cdot \cdot \cdot \quad (22)$$

We introduce the following variables into the integrand:

$$x = \frac{h\nu}{kT}, \quad \frac{h\nu_0}{k} = \Theta, \quad x_0 = \frac{h\nu_0}{kT} = \frac{\Theta}{T}. \quad \cdot \cdot \cdot \cdot \cdot \quad (23)$$

Thus  $\nu_0$  enters into the characteristic temperature in place of the characteristic frequency  $\nu$  of the preceding section. We then have

$$u = 9RT\Psi\left(\frac{\Theta}{T}\right), \quad \text{where } \Psi(x) = \frac{1}{x^3} \int_0^x \frac{x^3 dx}{e^x - 1}. \quad (24)$$

For *small* values of  $x$ , i.e. *high* temperatures, development of the denominator in a power series yields

$$\begin{aligned} \Psi(x) &= \frac{1}{x^3} \int_0^x \frac{x^3 dx}{x + \frac{x^2}{2} + \frac{x^3}{6} + \dots} \\ &= \frac{1}{x^3} \int_0^x x^2 \left( 1 - \frac{x}{2} + \frac{x^2}{12} - \dots + \dots \right) dx \\ &= \frac{1}{3} - \frac{x}{8} + \frac{x^2}{60} - \dots + \dots, \quad \cdot \cdot \cdot \cdot \cdot \quad (25) \end{aligned}$$

\* In equation (9) we were dealing with  $L = R/k$  oscillators.

and 
$$u = 3RT - \frac{9R\Theta}{8} + \frac{3}{20} \frac{R\Theta^2}{T} - \dots + \dots \quad (26)$$

If the series is terminated with the first term, differentiation again gives the specific heat in harmony with the Law of Dulong and Petit. The higher terms thus give the drop in the specific heat for smaller temperatures.

For the lower portion of the curve  $c_v = f(T)$ , i.e. for large values of  $x$ , we can write  $\infty$  as the upper limit, since the high values of  $x$  contribute very little to the integral on account of the occurrence of  $e^x$  in the denominator. The power series development of  $x^3(e^x - 1)^{-1}$  yields

$$x^3(e^x - 1)^{-1} = x^3e^{-x}(1 - e^{-x})^{-1} = x^3e^{-x} + x^3e^{-2x} + x^3e^{-3x} + \dots,$$

and term by term integration between 0 and  $\infty$  gives the series \*

$$\Psi(x) = \frac{6}{x^3} \left( 1 + \frac{1}{2^4} + \frac{1}{3^4} + \frac{1}{4^4} + \dots \right).$$

The sum of the series  $\Sigma(1/n^4)$  is  $\pi^4/90 = 1.0825$ . Finally, in the neighbourhood of the zero point, we have

$$u = 58.45 RT \cdot \frac{T^3}{\Theta^3}, \quad \dots \quad (27)$$

and for  $c_v$  near  $T = 0$ ,

$$c_v = 233.85 R \frac{T^3}{\Theta^3}, \quad \dots \quad (28)$$

This is the empirically well substantiated  $T^3$  law for the specific heat of a solid.

For further details, consult P. Debye, *Ann. der Physik*, **39**, p. 789 (1912).

\* Since

$$\begin{aligned} \int_0^\infty x^3 e^{-nx} dx &= -\frac{1}{n} \left[ x^3 e^{-nx} \right]_0^\infty + \frac{1}{n} \int_0^\infty 3x^2 e^{-nx} dx = -\frac{3}{n^2} \left[ x^2 e^{-nx} \right]_0^\infty \\ &+ \frac{3}{n^2} \int_0^\infty 2xe^{-nx} dx = -\frac{6}{n^3} \left[ x e^{-nx} \right]_0^\infty + \frac{6}{n^3} \int_0^\infty e^{-nx} dx = \frac{6}{n^4}. \end{aligned}$$

## CHAPTER XXXVII

### THE THEORY OF THERMAL RADIATION

#### 1. Analogy between a Gas and Radiation filling an Enclosure. Radiation Pressure.

The experiences of everyday life show that a hot body radiates light, i.e. electromagnetic waves. The problem of calculating the total amount of radiation emitted by unit area of a surface per second, and of determining its spectral distribution, is thus as much a matter for the Theory of Heat as for Electromagnetism.

First, we must ask whether the radiation of a hot body depends only upon its own properties and not also upon those of the environment. At first sight it might appear that, in analogy with thermal conduction, the radiation sent to a cool neighbouring body is more intense than that sent to a warmer body. An important advance in the solution of the problem formulated above was made by Prévost when he recognized that the rate of radiation is independent of the surroundings, and that the apparent dependence upon the environment is due to the fact that the cooler body in turn sends energy back to the radiator, but at a smaller rate than does a hot body, so that the side of the radiator which is presented to the colder body cools off more rapidly. In this manner the number of variables upon which the radiation of a hot body depends is considerably reduced. Further, we shall find that the radiation from any body may be computed if we know that of a perfectly black body, and if we know the absorption spectrum of the given body.

Finally, the problem of calculating the radiation from a perfectly black body is solved if we are able to compute the distribution of energy in a space bounded by walls which are impermeable to heat. In the Theory of Radiation such a space is usually called a *cavity* or *enclosure*. Let the walls be at a uniform temperature  $T$ . Such a cavity filled with radiation (the whole space is continually being traversed by electromagnetic waves radiated from the walls) has many points of similarity to a vessel filled with a gas. It is true that there is one point of difference: the density of a gas enclosed in a volume  $V$  may be changed at will, so that we have the density (or pressure) as a free variable, in addition to the temperature. On the other hand, the



energy density of the confined radiation is already completely fixed by the temperature. But just as the atoms of a gas exert pressure on the walls, so likewise does the confined radiation. This *radiation pressure*, whose magnitude will be calculated below, is very small under conditions available in the laboratory, but may attain values, in the interior of the stars, comparable with the gas pressure.

We know from experience that the spectrum of a glowing "black" body is continuous, i.e. all frequencies are present in the radiation. The problem of calculating the distribution of energy among the several frequencies is very similar to that of determining the distribution of the energy of a gas among the various molecular velocities. The superficial similarity between the experimentally determined energy distribution curve and the Maxwell distribution curve (fig. 1, p. 592) was recognized from the very first. No atom of a gas necessarily possesses exactly any selected velocity; we can only specify the number of atoms whose velocities fall within a given velocity interval  $\Delta c$ . In the same way, there is not a given amount of radiant energy corresponding to a single frequency  $\nu$ , but only an energy interval  $\Delta U$  corresponding to a prescribed range  $\Delta \nu$ , i.e. it is only the quotient  $\Delta U / \Delta \nu$  which is finite. To prevent misunderstandings, which occasionally appear in the literature, it must be pointed out that these considerations hold only for the continuous spectra resulting from *temperature radiation*. The statement occasionally made that there can be no sharp spectral lines, i.e. finite quantity of energy at a single frequency  $\nu$ , on account of thermodynamic reasons, is erroneous. If the excitation is not thermal, there is no reason for believing that a sharp line cannot be emitted. The fact that even in this case the line will really have a finite (but small) width is due to other causes.

Before attacking the actual problem of computing the distribution of energy in a radiation-filled enclosure, we shall deduce the value of the radiation pressure from the electrical field equations. We consider only the simplest case, that of a plane wave incident normally upon the plane surface of a perfect conductor. The ideal case  $\sigma = \infty$  implies a perfectly reflecting surface. An electric field cannot exist within a perfect conductor, for if a field did exist, it would cause infinite currents to flow; we may suppose, however, for convenience of calculation that  $\sigma$  is merely very great, not infinite, in which case there will be currents very near the surface of the conductor. Allow the incident wave to travel in the direction of the positive  $x$ -axis, and let the surface of the conductor be the  $yz$ -plane. The plane of vibration of  $\mathbf{E}$ , which is still undetermined, is taken to be the  $xy$ -plane. The equation of the incident wave is then

$$E_y = A e^{i\omega(t - x/c)} \quad . \quad . \quad . \quad . \quad . \quad (1)$$

The boundary condition concerning the continuity of the tangential

components of  $\mathbf{E}$  requires that the electric field strength be always zero for  $x = 0$ . (For perpendicular incidence there exists no normal component of  $\mathbf{E}$  whose discontinuity would give rise to a surface charge. There is such a component, however, for oblique incidence.) The boundary conditions are satisfied by adding a reflected wave of equal amplitude whose phase differs from that of the original wave by  $\pi$ ,

$$E_y' = -Ae^{i\omega(t+x/c)} \quad \dots \quad (2)$$

The perfect reflection of the mirror is represented by the equality of  $|E_y'|$  and  $|E_y|$ . Exactly as on p. 330, the corresponding magnetic vectors are determined from Maxwell's equations:

$$H_z = Ae^{i\omega(t-x/c)}, \quad H_z' = +Ae^{i\omega(t+x/c)}, \quad H_y = H_y' = H_x = H_x' = 0. \quad (3)$$

This indicates that the magnetic field does not vanish at the surface of the conductor, inasmuch as  $H_z$  and  $H_z'$  have the same sign. Since currents are flowing very close to the surface of the conductor, the magnetic field will exert a force on the conductor. By the results of p. 310, the force on a volume element  $d\tau$  traversed by a current whose density is  $\mathbf{i}$  is given by

$$d\mathbf{F} = \frac{[\mathbf{i}\mathbf{H}]}{c} d\tau. \quad \dots \quad (4)$$

Now the vector  $\mathbf{i}$  may be expressed in terms of curl  $\mathbf{H}$ . Since  $\mathbf{E}$  and  $\partial\mathbf{E}/\partial t$  vanish within the conductor, the first field equation yields

$$\mathbf{i} = \frac{c}{4\pi} \text{curl } \mathbf{H}. \quad \dots \quad (5)$$

Since only  $H_z$  differs from zero, curl  $\mathbf{H}$  has only a  $y$ -component, and we have

$$i_y = -\frac{c}{4\pi} \frac{\partial H_z}{\partial x}. \quad \dots \quad (5')$$

Substituting this in equation (4), there remains only an  $x$ -component of the force:

$$dF_x = -\frac{1}{4\pi} H_z \frac{\partial H_z}{\partial x} d\tau. \quad \dots \quad (6)$$

Take as a volume element a cylinder of unit cross-section and altitude  $dx$ . To find the total force acting on this cylinder, we integrate from 0 to  $\infty$ :

$$F_x = -\frac{1}{4\pi} \int_0^\infty H_z \frac{\partial H_z}{\partial x} dx = -\frac{1}{8\pi} [H^2]_0^\infty. \quad \dots \quad (7)$$

\* We write  $d\mathbf{F}$  for the force on the element instead of  $\mathbf{F}$ , as on p. 310, since we shall integrate over a finite volume later.



Since even for finite conductivity of the metal the wave penetrates but a very small distance, the magnetic field strength becomes zero for  $x = \infty$ , while, according to (3), it becomes  $|2Ae^{i\omega t}| = 2|H|$  for  $x = 0$ . This is the force acting on unit area of the surface; we may therefore call it the *pressure* of the radiation. Since for a plane wave in a vacuum

$$|E| = |H| \quad \text{and} \quad |S| = \frac{c}{4\pi} |E| |H|,$$

this pressure is, for normal incidence,

$$p_n = \frac{1}{2\pi} |E| |H| = \frac{2|S|}{c}. \quad . . . . (8)$$

In this formula  $|E|$ ,  $|H|$  and  $|S|$  signify the instantaneous values of the periodically varying quantities. What we observe, however, is the average in time,  $\bar{p}_n$ . This is obtained by inserting the time average of the Poynting vector on the right side of the equation. According to equation (60) (p. 341), this is equal to half the maximum amplitude, so that we have

$$\bar{p}_n = 2 \frac{\bar{S}}{c}. \quad . . . . (8')$$

Thus the pressure of radiation is proportional to the *normally incident* energy flux  $S$ . For oblique incidence we have, in addition, an electrical force acting on the surface charge. The final formula has  $\cos^2 \theta$  as an additional factor, where  $\theta$  is the angle of incidence.

The existence of such a pressure was first demonstrated experimentally and measured by Lébedew (1900), and in an extensive series of experiments conducted by Nichols and Hull (1903). The pressure found is in close agreement with the theoretical value.

## 2. The Connexion between Energy Density and Surface Brightness.

As already mentioned, the surface brightness is the radiation quantity which is experimentally measurable, while it is the energy density which is calculated by the theory. We therefore begin by finding the connexion between these two quantities. Consider an enclosure bounded by walls which are perfect heat insulators. Assume that bodies of any kind may be inserted into this space. Assume further that the dimensions of these bodies are large compared with the wave-length of the radiation, so that diffraction and scattering need not be taken into consideration. As a matter of fact, we wish to consider even differential segments large compared with the wave-length. Likewise, we wish to assume that the time intervals and their differentials are large compared with the period of the waves, so that the mean values of the radiation quantities are sufficiently defined. On account of the finite velocity of propagation of the electromagnetic



waves streaming through the space, there is a definite electromagnetic energy  $du$  in each element of volume. The limit  $du/d\tau$  is called the energy density  $\rho$ . This is spectrally analysed, so that the energy density  $\rho_\nu d\nu$  corresponds to the frequency interval  $d\nu$ . The task of the theory is to determine  $\rho_\nu$  for each interval as a function of the temperature of the enclosure. The *surface brightness* is defined as the amount of energy within the frequency interval  $d\nu$  radiated perpendicularly from an element of surface, per second per square centimetre, within unit solid angle. This quantity is represented by  $2K_\nu d\nu$ ; the factor 2 is present because the radiation can be resolved in two mutually perpendicular planes of polarization. Thus  $K_\nu$  is the surface brightness for linearly polarized radiation. Since we are considering isotropic bodies only, the surface brightness for unpolarized radiation is twice this quantity. The introduction of an angle of divergence (solid angle) of the pencil of rays is necessary on account of the fact that a source of light of finite surface brightness can produce no exactly parallel pencil. In order to obtain such a pencil one must have a perfect point source at the focus of a lens. If the beam coming from the lens is to have a finite intensity, the surface brightness of the point (i.e. extensionless) source must be infinite. If the normal to the surface which is emitting the radiation makes an angle  $\theta$  with the axis of the elementary cone, then—as may be seen from fig. 1—the energy flux per square centimetre is reduced to  $\cos \theta$  times the value for normal emergence, for at oblique emergence the surface concerned in the radiation is increased by the factor  $1/\cos \theta$ . This is known as Lambert's Law. It is a consequence of this law, for example, that a glowing cylinder appears to be a uniformly illuminated rectangle and the sun gives the impression of being a uniformly bright disc. The portions of the surface near the rim are foreshortened, but radiate correspondingly less in the oblique direction. Lambert's Law holds only for "black" thermal radiation, and is no longer valid, for example, for the X-radiation produced by the impact of electrons upon the anticathode; in this case it is possible to obtain increased intensity by observing obliquely.

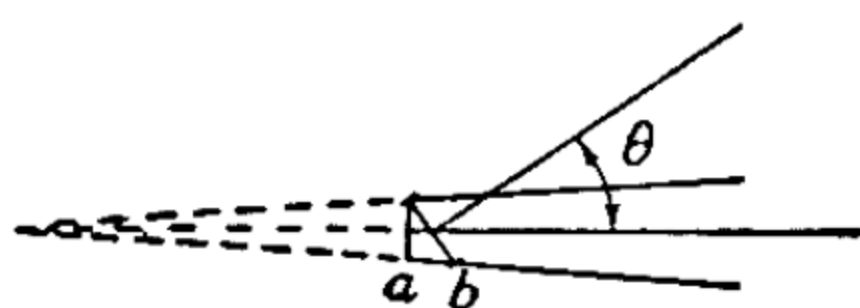


Fig. 1

Accordingly we have for the energy within the frequency range  $d\nu$  radiated in the time  $dt$  by the surface  $d\sigma$  in a solid angle

$$d\Omega = \sin \theta d\theta d\phi,$$

whose axis forms an angle  $\theta$  with the normal to the surface

$$d^6u = d^5u_\nu d\nu = d\nu dt d\sigma \cos \theta d\Omega \cdot 2K_\nu \quad \dots \quad (9)$$

As long as we consider only isotropic bodies,  $K_\nu$  is independent of the

direction, and we obtain the radiation within  $d\nu$  emitted by  $d\sigma$  in the time  $dt$  by integration over the hemisphere:

$$d^4u = d^3u_\nu d\nu = 2d\nu dt d\sigma K_\nu \int_0^{\pi/2} \int_0^{2\pi} \cos\theta \sin\theta d\theta d\phi \\ = 2\pi dt d\sigma K_\nu d\nu. \quad (10)$$

For the total radiation we integrate over the entire spectrum:

$$d^3u = 2\pi dt d\sigma \int_0^\infty K_\nu d\nu. \quad (11)$$

If, now, we consider a volume element  $d\tau$  within a medium which is being permeated by radiation, this element will be traversed by a large number of pencils of rays originating outside it. Consider a definite portion of the neighbourhood of  $d\tau$  bounded by a sphere of radius  $r$  whose centre is at the centre of  $d\tau$  (fig. 2). Resolve the radiation sent to  $d\tau$  from a *very small* element of surface of this sphere into elementary conical pencils. The portions of these cones within  $d\tau$  may be considered to be small cylinders. If the cross-section of such a cylinder is  $dS$ , the solid angle  $d\Omega$

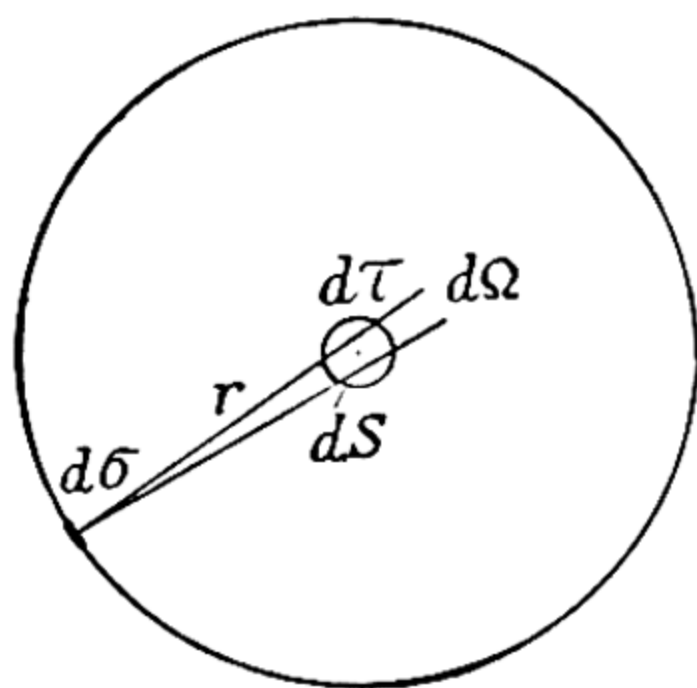


Fig. 2

of the small cone is  $dS/r^2$ , and the amount of energy radiated by  $d\sigma$  into this cone per second is

$$\frac{d^6u}{dt} = 2d\nu d\sigma K_\nu \frac{dS}{r^2}. \quad (12)$$

This may be looked upon as the part of the energy flux within the elementary cone originating at  $d\sigma$ . The density of the energy flow is then obtained by dividing by  $dS$ . Now for any flow, the current density of any quantity is equal to the product of the density of the quantity by the velocity of flow (see, for example, p. 418). The latter is  $c/n$ , where  $n$  is the index of refraction of the medium. Thus the energy density in that part of the volume element pierced by the given cone which is due to the radiation from  $d\sigma$  is given by

$$d^3\rho = d^2\rho_\nu d\nu = \frac{2d\nu d\sigma}{r^2 \cdot c/n} K_\nu. \quad (13)$$

This value holds for all parts of the element of volume  $d\tau$ , since the other cones which penetrate  $d\tau$  differ infinitesimally in direction; hence  $d^3\rho$  gives the energy density in  $d\tau$  due to the radiation from  $d\sigma$ . We then have by integration over all surface elements of the sphere

$$d\rho = \rho_\nu d\nu = \frac{8\pi K_\nu}{c/n} d\nu, \quad (14)$$



whence

$$\rho_\nu = \frac{8\pi K_\nu}{c/n} \quad . \quad . \quad . \quad . \quad . \quad . \quad (15)$$

### 3. The Connexion between Emission and Absorption in Thermodynamic Equilibrium. Kirchhoff's Law.

Consider a uniform medium within the enclosure. At equilibrium, every element of volume of this medium will radiate a certain amount of energy each second, i.e. will convert a given amount of heat energy into radiation, and will absorb an equal amount of radiation delivered by the environment, converting it into heat. With regard to the process of emission and absorption, it must be remembered this takes place in the atoms of the substance, so that a certain volume is necessary for the emission of a given quantity of energy. A mathematical surface contains no atoms, and so can emit no stated amount of energy. It is true that we often speak of the emissive power of a surface; in such cases we mean the total energy which streams through the surface from the interior of the body. Hence we define the *emissivity*, or *coefficient of emission*  $\epsilon_\nu d\nu$  as the energy of linearly polarized radiation within the frequency interval  $d\nu$ , referred to unit volume and unit solid angle, emitted in unit time. Then the total amount of unpolarized energy in the range  $d\nu$  emitted in the solid angle  $4\pi$  in the time  $dt$  is

$$d^5u_e = 8\pi d\nu dt d\tau \epsilon_\nu \quad . \quad . \quad . \quad . \quad . \quad (16)$$

The absorption coefficient  $\alpha_\nu$  for frequencies in the range  $\nu$  to  $\nu + d\nu$  is defined in the manner which is usual in optics as the fraction of the incident radiation absorbed per second in a thin layer, referred to unit thickness. We again surround the element of volume by a sphere of radius  $r$ . The rays emanating from an element  $d\sigma$  of the sphere and penetrating the element of volume fill a solid angle which again may be subdivided into elementary cones; the portions of these cones within the volume element may be considered cylinders, as before. If these cylinders are of altitude  $dl$ , and of cross-section  $dS$ , then the total amount of energy absorbed in the time  $dt$  from the unpolarized pencil coming from  $d\sigma$  is given by

$$d^7u_a = 2dt d\nu \Sigma \frac{dS}{r^2} dl \alpha_\nu d\sigma K_\nu \quad . \quad . \quad . \quad . \quad (17)$$

Now  $\Sigma dS dl$  is the volume of the element  $d\tau$ . Integration over all surface elements of the sphere yields the following expression for the total amount of energy absorbed from the environment:

$$d^5u_a = 8\pi dt d\nu d\tau \alpha_\nu K_\nu \quad . \quad . \quad . \quad . \quad (18)$$



If we set the emitted energy equal to that absorbed, we obtain simply

$$K_\nu = \frac{\epsilon_\nu}{\alpha_\nu}. \quad . . . . . (19)$$

*For any uniform medium in thermodynamic equilibrium, the ratio of the emission coefficient to the absorption coefficient for any spectral region is equal to the surface brightness.* This statement is often referred to as Kirchhoff's Law, but the latter refers to two other quantities, especially in its experimental applications, as we shall see immediately.

If the enclosure is filled with a medium which is perfectly transparent to all rays, e.g. a vacuum (or even air, as long as the system is at temperatures available in the laboratory), then while it is true that  $K_\nu$  is defined by the energy passing through a surface, the fraction  $\epsilon_\nu/\alpha_\nu$  is of the form  $0/0$ . But the enclosure is bounded by walls of some kind, from which the radiation comes. We therefore investigate the radiation exchange at the boundary. The amount of energy passing through an area  $d\sigma$  of the wall within a solid angle  $d\Omega$  whose axis makes an angle  $\theta$  with the normal, in the time  $dt$ , is

$$d^6u = 2d\nu dt d\sigma d\Omega \cos \theta K_\nu. \quad . . . . . (20)$$

Since there is thermodynamic equilibrium, the same amount of energy must pass through this same surface element in the opposite direction. Of this energy flow, only a part determined by the emissive power  $E_\nu$  will come from the interior of the material constituting the wall, the rest being radiation which is reflected by the wall. The fraction of the incident radiation which is absorbed in the substance of the wall is called the *absorptive power*  $A_\nu$  for the frequency range  $\nu$  to  $\nu + d\nu$ . The part  $1 - A_\nu$  is reflected at the surface of the wall. The energy equation at this surface is then

$$2d\nu dt d\sigma d\Omega \cos \theta K_\nu = (1 - A_\nu) 2d\nu dt d\sigma d\Omega \cos \theta K_\nu + 2d\nu dt d\sigma d\Omega \cos \theta E_\nu, \quad . (21)$$

whence 
$$K_\nu = \frac{E_\nu}{A_\nu}. \quad . . . . . (22)$$

This is Kirchhoff's Law. *The ratio of the emissive power to the absorptive power of a body is equal to the surface brightness in the adjacent diathermanous medium.* The difference between the concepts of emissive and absorptive power introduced here and the emission and absorption coefficients is that the former depend upon the thickness of the body. Even if the absorption coefficient is small, a sufficiently deep layer can destroy practically all the incident energy. A body for which  $A_\nu = 1$  is called a perfectly black body. For example, a sphere of gas—representing, say, a star—may be considered a black body on

account of the extremely great depth of material, although the absorption *coefficient* of the gases is very small.

The fact that the emissive power of a black body is equal to the surface brightness in the adjacent medium makes possible the construction of a black body in a simple way. While black substances like *platinum black* absorb visible light completely, they are no longer "black" for longer waves, so that the use of such substances as black bodies may lead to deception. According to what has been said, a black body free from these objections is obtained by taking a large enclosure in the wall of which there is a small hole whose diameter must be so small compared with the size of the container that the radiation passing outward through the hole is negligible; its surface brightness is that which obtains in the diathermanous medium. The surface of this hole then represents that of a perfect black body. The effectiveness of this device may be seen by cutting a small hole in the side of a box made of black paper. The hole appears to be much blacker than the paper. In fact, the emissive power of a black body is experimentally determined by measuring the surface brightness of a small opening in the side of a uniformly heated furnace.

According to Kirchhoff's Law the thermal emission of any other body is obtained, for any spectral region, by multiplying the black body value by the spectroscopically determined absorptive power. The emissivity of a non-black body must be less than that of a black body for all colours, and can attain the black body value only in the regions of strong selective absorption. In instances where the emissivity exceeds that of a black body, e.g. for a gas mantle, the excitation is not purely thermal. Here, for example, the gas in the region of combustion is by no means in thermal equilibrium, and the glow is due in part to chemiluminescence. The same is true of "cold" flames, e.g. the zone of reaction of a sodium vapour jet in chlorine, which gives the yellow sodium line with considerable intensity at temperatures as low as 500° C.

#### 4. Planck's Law of Radiation.

The relation between the emissive power of a black body and the energy density of radiation in an enclosure having been made clear in the last two sections, it remains only to calculate the energy density according to the methods of statistics. All the necessary computational work has already been done in Chap. XXXVI, §3 (p. 608). Consider the enclosure to be a cube of edge  $l$ . Since the electromagnetic waves traversing this space are pure transverse waves, the number of possible frequencies in the range  $\nu$  to  $\nu + d\nu$  is again given by

$$dZ = \frac{8\pi V \nu^2}{c^3} d\nu. \quad . \quad . \quad . \quad . \quad . \quad (23)$$



Each wave is statistically equivalent to an oscillator, and since, by equation (9) (p. 606), the energy of a system consisting of  $N$  oscillators of frequency  $\nu$  is

$$U = \frac{N h \nu}{e^{h\nu/kT} - 1}, \quad \cdot \cdot \cdot \cdot \cdot \cdot (24)$$

the energy density within the interval  $d\nu$ , per cubic centimetre, is

$$\rho_\nu d\nu = \frac{8\pi h \nu^3}{c^3(e^{h\nu/kT} - 1)} d\nu. \quad \cdot \cdot \cdot \cdot \cdot (25)$$

Hence for unpolarized radiation the emissive power of a black body, in the range  $d\nu$ , becomes

$$2E_\nu d\nu = 2K_\nu d\nu = \frac{2h\nu^3}{c^2(e^{h\nu/kT} - 1)} d\nu. \quad \cdot \cdot \cdot (26)$$

This is the celebrated Planck Radiation Law. The curve  $E_\nu(\nu)$  has the typical form of a Maxwell velocity distribution curve (cf. fig. 1, p. 592): first, a steep increase to a maximum, then an exponential decline toward higher frequencies. This exponential decay is characteristic of the Quantum Theory. The Rayleigh-Jeans Law, based on the equipartition of energy, requires the number of waves in a given interval to increase as  $\nu^2$ , so that it would yield a continued increase of energy toward high frequencies, i.e. toward the ultra-violet, which is contrary to experience.

If it is desired to express Planck's Law in terms of the wave-length, it must be remembered that for a spectral interval expressed in  $d\nu$  and again in  $d\lambda$ ,

$$E_\nu d\nu = E_\lambda d\lambda,$$

and that

$$|d\nu| = \frac{c}{\lambda^2} |d\lambda|.$$

The radiation law (26) then becomes

$$2E_\lambda d\lambda = \frac{2hc^2}{\lambda^5} \cdot \frac{1}{e^{hc/kT\lambda} - 1} d\lambda. \quad \cdot \cdot \cdot \cdot (27)$$

Fig. 3 is a graphical representation of the content of this equation.

The Law of Planck contains all other laws of radiation. The Wien Displacement Law, which was derived from quite different considerations by W. Wien before the Planck law was known, relates to the position of the maximum of the energy distribution curve. It is especially simple when expressed in terms of  $\lambda$ : differentiating the right member of (27) and setting the result equal to zero, we have

$$-\frac{10hc^2}{\lambda^6(e^{hc/kT\lambda} - 1)} + \frac{2hc^2 e^{hc/kT\lambda}}{\lambda^5(e^{hc/kT\lambda} - 1)^2} \cdot \frac{hc}{kT\lambda^2} = 0.$$



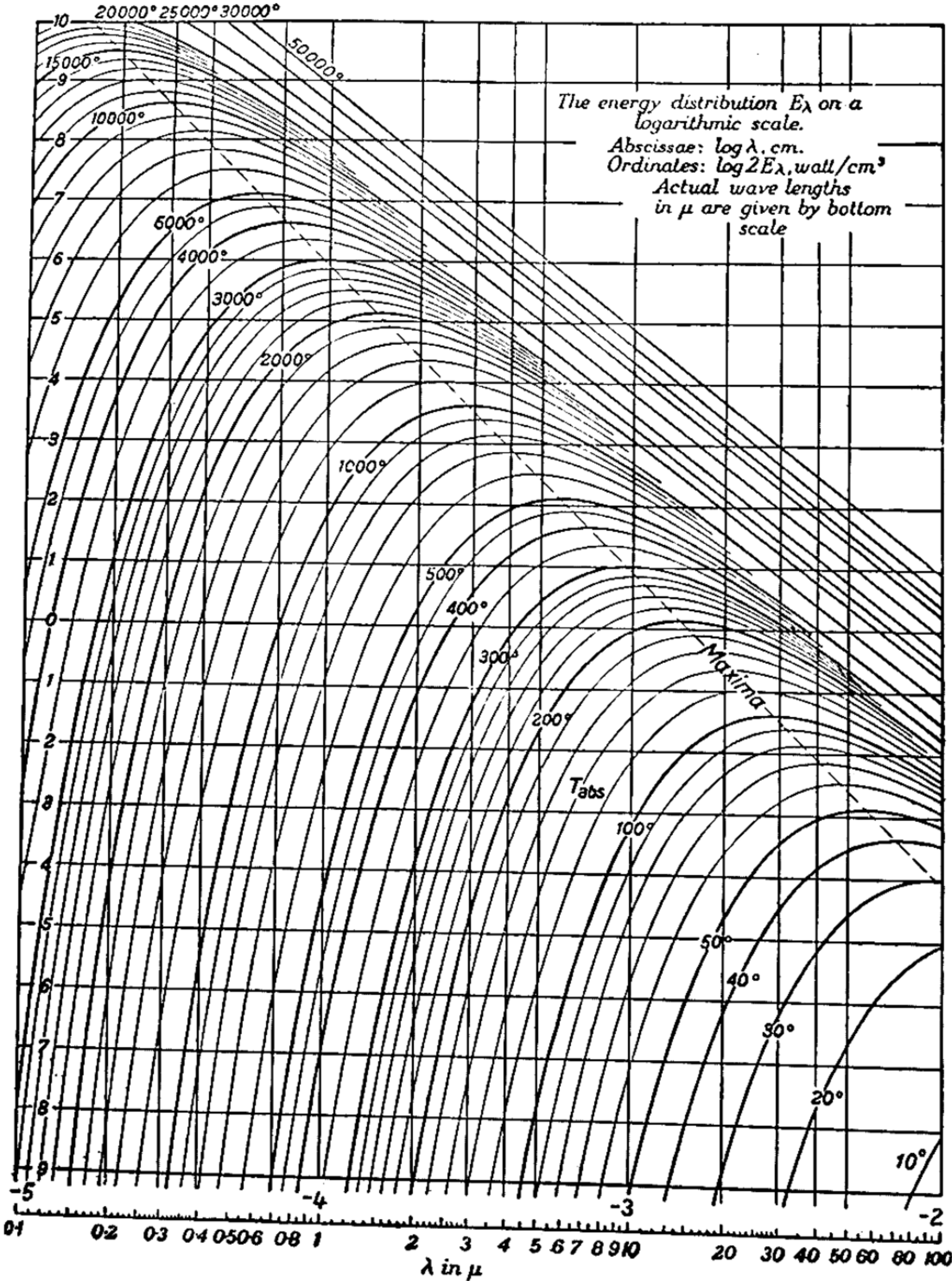


Fig. 3

Eliminating the common factors and introducing the variable  $x = hc/kT\lambda$ , this becomes the transcendental equation

$$\frac{xe^x}{e^x - 1} = 5.$$

We see at once that there must be a root in the neighbourhood of 5. Applying the usual methods of approximation, the exact value is found to be 4.965. This is the only real root, as may be seen readily by drawing the curve. Hence we have

$$\lambda_{\max.} = \frac{hc}{4.965kT},$$

or 
$$\lambda_{\max.} T = \frac{hc}{4.965k} = b = 0.2896 \text{ cm. deg.} \quad . \quad . \quad (28)$$

This is Wien's Displacement Law, which states that *as the temperature of a black body is increased, the position of maximum emission moves in the direction of shorter waves in such way that the product  $\lambda_{\max.} T$  remains constant.*

The total radiation is obtained by integrating over all frequencies:

$$2E = 2 \int_0^\infty E_\nu d\nu = \frac{2h}{c^2} \int_0^\infty \frac{\nu^3 d\nu}{e^{h\nu/kT} - 1} = \frac{2k^4 T^4}{c^2 h^3} \int_0^\infty \frac{x^3 dx}{e^x - 1}.$$

This integral has been evaluated on p. 612, where its value was found to be  $\pi^4/15$ . Thus we have

$$2E = \frac{2\pi^4 k^4}{15c^2 h^3} T^4 = \sigma' T^4. \quad . \quad . \quad . \quad (29)$$

This is the Stefan-Boltzmann Law, according to which *the total energy emitted by a black body varies as the fourth power of the absolute temperature.* Usually, however, we do not give the energy radiated normally into a unit solid angle, but rather the total energy radiated on one side of a plane, i.e. into a solid angle  $2\pi$ . The constant of proportionality  $\sigma$  so defined then has the value  $\pi\sigma'$ , according to equation (10) (p. 618), so that

$$\sigma = \frac{2\pi^5 k^4}{15c^2 h^3} = 5.68 \times 10^{-5} \frac{\text{erg.}}{\text{cm.}^2 \text{ deg.}^4 \text{ sec.}}$$

*Ex. 130.* Using a cyclic process, deduce the Stefan-Boltzmann Law, leaving the factor of proportionality undetermined. The procedure resembles that by which the vapour pressure formula was obtained (on p. 534 *et seq.*). Show also that the pressure of radiation amounts to  $p = \rho/3$ .

## CHAPTER XXXVIII

### THE BOSE-EINSTEIN AND THE FERMI-DIRAC STATISTICS

#### 1. Preliminary Remarks concerning the Duality of Waves and Corpuscles.

In the preceding chapter we obtained the law of distribution of energy in an enclosure filled with radiation by counting the number of electromagnetic waves which could exist there. Now there are numerous phenomena where radiation is converted into energy which is connected with matter—kinetic energy of electrons, chemical energy, &c. In such instances light behaves as though it consisted of discrete corpuscular elements of energy, *light quanta* of energy  $h\nu$ . On the other hand, the corpuscular idea is not capable of including all properties of radiation, for an immense number of interference phenomena demonstrate the wave-like character of light. In Part VII of this work we shall bring out how the wave concept and the particle concept are two aspects of the same thing, and how the facts may be represented only by a combination of both ideas. We shall arrive at the conclusion that the light quanta are guided by the wave field, and that it is not possible to specify the positions of the quanta inside a wave. In view of the analogy between an enclosure filled with radiation and one filled with gas, it is to be expected that a corpuscular theory will also yield the radiation law correctly. When, however, we calculate the energy distribution of the quanta according to the Boltzmann statistics, we obtain the incorrect radiation law in which the energy is displaced into the short wave end of the spectrum. The underlying reason for the failure of this method is that we assume the light quanta to be independent of one another when applying the Boltzmann method. This independence certainly does not exist; if two coherent waves (bundles of quanta) interfere, this implies a mutual influence between the quanta, leading to a decrease in intensity at certain points and an increase at others. At any rate, a displacement of quanta occurs.\* The correct radiation law is obtained by a new statistical method introduced by Bose.

\* It is well to remember that there is no actual "destruction" of light in interference phenomena.



Further, a number of recent experiments indicate that it is also incorrect to assume the independence of particles even when dealing with electrons and with entire atoms. There are guiding waves for these particles too, and these waves exhibit the properties of diffraction and interference. These ideas will be discussed in greater detail in Chap. XL (p. 687). At this point we wish to say only that the "statistics of direct intuition"—as we may call the Boltzmann statistics—is not applicable. We may then build up a new form of statistics either in terms of waves or in terms of corpuscles. The wave form would have the advantage that its deductions are more illuminating, once we assume the existence of the guiding waves. Nevertheless, the results are often needed in the corpuscular form. It is preferable, therefore, to employ the statistics in terms of corpuscles; this is the form deduced by Bose, by means of a happy artifice, for light quanta. For ordinary gases the difference between this and the classical statistics is so small that it has not yet been found possible to distinguish between the two experimentally. For an electron gas, however, the situation is such that for densities of the order of that attributed to the electron gas inside metals the results according to the two views are totally different. For the case of electrons, however, the Bose statistics must be supplemented so as to conform with a certain general principle of atomic structure. In this way we arrive at a new form of statistics due to Fermi—a form which is capable of solving many of the problems relating to electrons in metals.

## 2. The Bose-Einstein Statistics.

The essential characteristic of the statistics devised for light quanta by Bose \* and applied to actual gases by Einstein † is that no distinction is made between individual atoms, even in the micro-states. In the case of light quanta it seems artificial to distinguish between particles; on the other hand, the calculation of probability as carried out in classical statistics seemed reasonable and natural, because we are accustomed to the concept of individual atoms. The phenomena of the diffraction of atoms (p. 695) compel us to give up the idea of individually distinguishing the atoms or molecules of a given substance also. The way in which this alters the computation of probabilities will be illustrated by a very simple example. Suppose we wish to distribute three atoms among three cells. A macro-state of classical statistics is determined by giving the number of atoms in the first, second and third cells. We may summarize the various distributions as follows:

\* S. N. Bose, *Zeitsch. f. Phys.* 26, p. 178 (1924).

† A. Einstein, *Berl. Ber.* p. 261 (1924). *ibid.* p. 18 (1925).

Cell \ Group	I	II	III
1	3 0 0	2 0 1 2 1 0	1
2	0 3 0	1 2 2 0 0 1	1
3	0 0 3	0 1 0 1 2 2	1
Boltzmann Probability	1 1 1	3 3 3 3 3 3	6

As long as we look upon the atoms as distinguishable one from another, the distributions of the first group can be realized in but one way, the second in three ways, and the third in six ways—for a distribution in which Atom 1 is in Cell 1 and Atom 2 is in Cell 2 is taken to be different from a distribution in which Atom 2 is in Cell 1 and Atom 1 is in Cell 2. In Bose's statistics no such distinction is made, so that each of these distributions has the same probability, and, according to Bose, a distribution of this kind represents only a *micro-state*. What, then, is a macro-state in the Bose statistics? A macro-state may be represented by various micro-states, and the number of these gives its probability. It is natural to discard the distinction between the *cells* and to characterize the macro-state by specifying how many cells are occupied by 0, by 1, by 2, &c. atoms. Hence in the Bose statistics the macro-states are the groups bearing Roman numerals in the above tabulation. Thus State I has the probability 3, State II the probability 6, State III the probability 1. It is to be noticed, however, that we are permitted to give up the distinction only between cells which are physically equivalent, e.g. those of equal energy. For instance, if we wish to compute the velocity distribution of the molecules of an ideal gas according to the new method, the zone of the phase space—to which a given interval of molecular energy corresponds—already contains so many cells that we may perform the above probability computation within it. The general formula for the probability of a macro-state defined in such a way that of a total of  $Z$  cells  $Z_0$  are occupied by 0 particles,  $Z_1$  by 1, . . .  $Z_n$  by  $n$ , is evidently obtained from the following scheme by permuting the numbered *cells* instead of the numbered *particles* as heretofore:

*Schematic Representation of a Micro-state in the  
Bose Statistics*

Cell number:	1	2	3	4	5	6	
Occupied by:	0	1	0	2	2	3	particles
Macro-state: $Z_0=2, Z_1=1, Z_2=2, Z_3=1, Z_n=0$ for $n>3$							

Should it happen that two cells are occupied by the same number



of particles, the interchange of their numbers will yield no new complexion, and hence the probability becomes

$$W = \frac{Z!}{\prod Z_n!}, \quad \cdot \cdot \cdot \cdot \cdot \cdot \cdot \quad (1)$$

where  $\Sigma Z_n$  is equal to  $Z$ , the total number of cells and  $\Sigma nZ_n$  is equal to  $N$ , the number of particles.

Let us apply the method to the calculation of the entropy and distribution of velocities in an ideal monatomic gas. Since each molecule has three position and three momentum co-ordinates, the size of each cell of phase space must be taken to be  $h^3$ , as in classical statistics. Since the energy does not depend on the position, we can split up the phase space into the product of momentum space by position space, just as we did on p. 590—i.e. we take the element of volume of momentum space, which alone concerns us in what follows, equal to

$$\Delta\tau_p = \frac{h^3}{V}, \quad \cdot \cdot \cdot \cdot \cdot \cdot \cdot \quad (2)$$

where  $V$  is the volume available to the molecules.\* The energy of a molecule is given by

$$u = \frac{1}{2m}(p_x^2 + p_y^2 + p_z^2) = \frac{1}{2m}p^2. \quad \cdot \cdot \cdot \quad (3)$$

Hence the value of the energy is constant over a sphere in the momentum space. If we divide this space into spherical shells of thickness  $\Delta p$ , there will be a large number of elementary cells of size  $h^3/V$  in each spherical shell; the Bose statistics can be applied to the distribution in the several shells, since the cells of each spherical layer possess equal energy. Distinguishing the shells by the subscript  $s$ , the probability of a state in which  $Z_{ns}$  cells of the  $s$ th shell each contain  $n$  molecules is given by

$$W_s = \frac{Z_s!}{\prod Z_{ns}!}. \quad \cdot \cdot \cdot \cdot \cdot \cdot \cdot \quad (4)$$

According to the Theory of Probability we obtain the complete probability of any distribution by taking the product of the separate

\* Since  $\Delta\tau = \Delta x\Delta y\Delta z\Delta p_x\Delta p_y\Delta p_z$ , where  $\Delta\tau$  is the volume element of phase space it may seem at first sight that we have to set  $\Delta\tau_p = h^3/\Delta V$ , but since the Quantum Theory prescribes only the volume, not the form, of the element, we can extend the domain of the position co-ordinates, upon which the energy does not depend, at the expense of the range of the momentum co-ordinates. The deeper reason for being able to proceed in this manner is to be found in Heisenberg's Principle of Uncertainty (cf. p. 690), according to which the product of the possible variations of a position and a momentum co-ordinate is equal to  $h$ . Thus, if we give the position co-ordinate a large range of variation, the momentum co-ordinate is, correspondingly, more sharply defined. Hence, since the energy does not depend on position in the present instance, we can take the entire volume as the range of the position co-ordinates.



probabilities. Assume the Boltzmann equation to hold for this quantity:

$$\begin{aligned} S &= k \log W = k \log \prod_s W_s \\ &= k \sum_s \log Z_s! - k \sum_s \sum_n \log Z_{ns}! \\ &= k \sum_s Z_s \log Z_s - k \sum_s Z_s - k \sum_s \sum_n Z_{ns} \log Z_{ns} + k \sum_s \sum_n Z_{ns}, \end{aligned}$$

and since  $\sum_n Z_{ns} = Z_s$ ,

$$S = k \sum_s Z_s \log Z_s - k \sum_s \sum_n Z_{ns} \log Z_{ns}. \quad . \quad . \quad . \quad (5)$$

For the condition of equilibrium,  $W$  must again be a maximum, with the auxiliary conditions

$$\sum_n Z_{ns} = Z_s = \text{const.}, \quad . \quad . \quad . \quad . \quad . \quad (6)$$

for the total number of cells in each spherical layer is, of course, fixed by the geometry alone. Equation (6) then amounts to a number of auxiliary conditions equal to the number of shells. Further, we must have

$$\sum_s \sum_n n Z_{ns} = N, \quad . \quad . \quad . \quad . \quad . \quad (7)$$

where  $N$  is the total number of molecules, and

$$\sum_s \sum_n n u_s Z_{ns} = U, \quad . \quad . \quad . \quad . \quad . \quad (8)$$

where  $U$  is the total energy and  $u_s$  is the energy of a molecule whose representative point lies in the  $s$ th cell. We again determine the extremal value of the entropy under the given auxiliary conditions by applying Lagrange's method of multipliers. In taking the variation of the number of particles in each cell it must be remembered that the total number of cells in each layer is fixed, i.e. the numerator of the fraction in equation (4) is a constant. Then, from (4) and (5),

$$0 = \sum_s \sum_n \delta Z_{ns} (\log Z_{ns} + 1) = 0. \quad . \quad . \quad . \quad . \quad (9)$$

The variation of the auxiliary conditions gives

$$\sum_n \delta Z_{ns} = 0 \quad . \quad . \quad . \quad . \quad . \quad (10)$$

for each shell; moreover, for a constant total number of particles and constant intrinsic energy,

$$\sum_s \sum_n n \delta Z_{ns} = 0, \quad . \quad . \quad . \quad . \quad . \quad (11)$$

and

$$\sum_s \sum_n n u_s \delta Z_{ns} = \delta U = 0. \quad . \quad . \quad . \quad . \quad (12)$$

We now multiply each of the equations of (10) by a coefficient which

we take to be  $-(\log a_s + 1)$ , and multiply (11) by  $\beta$  and (12) by  $\gamma$ . Adding these to the main equation (9) and setting the coefficients of  $\delta Z_{ns}$  equal to zero yields

$$\log Z_{ns} - \log a_s + n\beta + nu_s\gamma = 0,$$

or 
$$Z_{ns} = a_s e^{-n(\beta + \gamma u_s)}. \quad \dots \quad (13)$$

The multipliers  $a_s$  may be determined by introducing the number of cells in each shell:

$$Z_s = \sum_n Z_{ns} = \sum_n a_s e^{-n(\beta + \gamma u_s)} = \frac{a_s}{1 - e^{-(\beta + \gamma u_s)}}. \quad \dots \quad (14)$$

The last form is the sum of the geometric progression. Then

$$a_s = Z_s(1 - e^{-(\beta + \gamma u_s)}). \quad \dots \quad (14')$$

For the number of particles in each shell—which is the goal of our computation—we have

$$\begin{aligned} N_s &= \sum_n n Z_{ns} = a_s \sum_n n e^{-n(\beta + \gamma u_s)} = -a_s \frac{\partial}{\partial(\beta + \gamma u_s)} \sum_n e^{-n(\beta + \gamma u_s)} \\ &= -a_s \frac{\partial}{\partial(\beta + \gamma u_s)} \left( \frac{1}{1 - e^{-(\beta + \gamma u_s)}} \right) = \frac{a_s e^{-(\beta + \gamma u_s)}}{\{1 - e^{-(\beta + \gamma u_s)}\}^2}. \end{aligned} \quad (15)$$

Inserting the value of  $a_s$ , this becomes

$$N_s = \frac{Z_s}{e^{(\beta + \gamma u_s)} - 1}. \quad \dots \quad (16)$$

It remains only to determine the multipliers  $\beta$  and  $\gamma$ . Since  $\gamma$  is multiplied by the energy of the cells, the total energy is a function of  $\gamma$  and, reciprocally,  $\gamma$  is a function of the total energy  $U$ . Inserting (16) in the entropy equation and taking the derivative  $(\partial S / \partial U)_v = 1/T$ , we obtain, exactly as on p. 588, the value  $1/kT$  for  $\gamma$ .\* In order to determine  $\beta$  we must turn to equation (7), which gives the total number of particles. The summation or integration, however, involves computational difficulties, so that we shall leave  $\beta$  undetermined. Setting  $e^\beta = B$ , we have the distribution formula

$$N_s = \frac{Z_s}{B e^{u_s/kT} - 1}. \quad \dots \quad (17)$$

For an ideal gas, the layers within which the energy may be considered constant are the spherical shells included between the radii  $p_s$  and  $p_s + dp_s$ . The equations connecting energy and momentum are

$$u_s = \frac{1}{2m} p_s^2, \quad \text{i.e. } p_s = \sqrt{2mu_s}, \quad dp_s = \sqrt{\frac{m}{2u_s}} du_s. \quad (18)$$

\* The corresponding multiplier was there denoted by  $\beta$ .

Hence a shell contains

$$dZ_s = \frac{4\pi\sqrt{2m^3u_s} du_s}{h^3/V} \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad (19)$$

cells of volume  $h^3/V$ , and, omitting the subscript  $s$ , the number of particles in the separate shells becomes

$$dN = \frac{4\pi V\sqrt{2m^3u} du}{h^3(Be^{u/kT} - 1)} \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad (20)$$

Finally, expressing  $u$  in terms of the velocity, we have the distribution formula

$$dN = \frac{4\pi Vm^3c^2 dc}{h^3(Be^{mc^2/2kT} - 1)} \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad (20')$$

If  $B \gg 1$ , we obtain

$$dN = \frac{4\pi Vm^3c^2 e^{-mc^2/2kT}}{h^3 B} dc, \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad (21)$$

while the classical statistics gave (p. 592)

$$dN = \frac{4Nc^2 e^{-mc^2/2kT}}{\sqrt{\pi} \left(\frac{2kT}{m}\right)^{3/2}} dc. \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad (22)$$

Thus when  $B$  is large, the Bose-Einstein statistics also leads to the Maxwell distribution. Equating the two expressions determines  $B$ :

$$B = \frac{V(2\pi mkT)^{3/2}}{Nh^3} \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad (23)$$

We are justified in equating the two expressions for the reason that integration over all velocities must yield the total number  $N$  of molecules in each case. The preceding equation shows that  $B$  is inversely proportional to  $N$ , so that our assumption of  $B \gg 1$  has greater validity—and hence the Bose distribution agrees more nearly with the Maxwell form—when  $N$  is small, i.e. when the gas is at low pressure. Actually, deviations are first to be expected at those densities and those low temperatures at which all gases begin to exhibit strong van der Waals departures from the ideal gas law. Up to the present time it has not been found possible to separate the two kinds of deviations to the extent that one can decide with certainty whether the gas obeys the Bose statistics or that of Boltzmann.

### 3. The Fermi-Dirac Statistics and its Application to Electrical Conductivity.

While the Bose statistics has not yet become of much practical importance on account of the difficulty of testing it experimentally it is, nevertheless, of great service in deriving a new form of statistics given



by Fermi and independently by Dirac.\* This method appears to be the appropriate means of treating the electron gas in metals. In the theory of the structure of the atom it is shown that a given combination of quantum numbers can appear but once in each atom. This law, which is of fundamental importance, is known as Pauli's Principle (cf. p. 679). Otherwise stated, this law says that a cell of phase space can be occupied by only one representative point. Fermi applied this principle to the totality of particles in a gas. Thus the numbers  $Z_{ns}$  vanish for all values of  $n$  except 0 and 1. If the computation is performed with this added hypothesis, all calculations for determining the most probable state remain unaltered. One must remember only that the summations for  $n$  are to be extended only from 0 to 1. The actual summations are not performed until the multipliers  $\alpha_s$  are determined. At this point the method departs from that of the Bose statistics. We have, at first, as for the Bose statistics,

$$Z_{ns} = \alpha_s e^{-n(\beta + \gamma u_s)}. \quad . \quad . \quad . \quad . \quad . \quad (24)$$

To determine the  $\alpha_s$  we sum over all values of  $n$ ; in this case the sum reduces to

$$Z_s = \alpha_s (1 + e^{-(\beta + \gamma u_s)}). \quad . \quad . \quad . \quad . \quad . \quad (25)$$

The number of particles in the shell of subscript  $s$  becomes

$$N_s = 1 \cdot Z_{1s} = \alpha_s e^{-(\beta + \gamma u_s)} = \frac{Z_s}{e^{(\beta + \gamma u_s)} + 1}. \quad . \quad . \quad (26)$$

As in § 2, (p. 630)  $\gamma$  is found to be  $1/kT$ , so that

$$N_s = \frac{Z_s}{B e^{u_s/kT} + 1}. \quad . \quad . \quad . \quad . \quad . \quad (27)$$

This result has a remarkable similarity to the Bose distribution function—the difference is only in the sign of the 1 in the denominator. This shows that the Fermi distribution, too, approaches the Maxwell-Boltzmann for large values of  $B$ , i.e. for low densities. Hence for ordinary gases we can no more expect to test the validity of the Fermi statistics than we could that of the Bose method. But the situation is entirely different for an electron gas of the type we are forced to assume exists within a metallic conductor (cf. p. 444 *et seq.*). If we assume that the number of conduction electrons is of the order of magnitude of the number of atoms, the density of the electron gas becomes so large that  $B$  is very small, especially since the small mass of the electron operates, according to equation (23), to affect the result in the same way. *At ordinary temperatures, assuming the Fermi*

\* E. Fermi, *Atti Lincei* (6), 3, p. 145 (1926); *Zeitschr. f. Physik.* 38, p. 902 (1926).  
P.A.M. Dirac, *Proc. Roy. Soc.*, 112A, p. 661 (1926).

statistics to be valid, the electron gas is completely degenerate, i.e. it obeys a totally different distribution law from that for an ideal gas.

We now wish to use the Fermi statistics to derive the most important property of the electron gas. First, we must double the number of cells belonging to a given energy range, as compared with the number for an arbitrary gas. The reason is this: it will be shown in Part VII that it is necessary to ascribe to the electron a magnetic moment of one Bohr magneton. According to the rules of the Quantum Theory, the axis of this moment can set itself either in the same or the opposite sense to a given field direction. For each of these positions we have the number deduced above for the number of cells of the translational momentum space. If we now pass to a very weak field, the magnetic contribution to the energy vanishes and the two orientations become

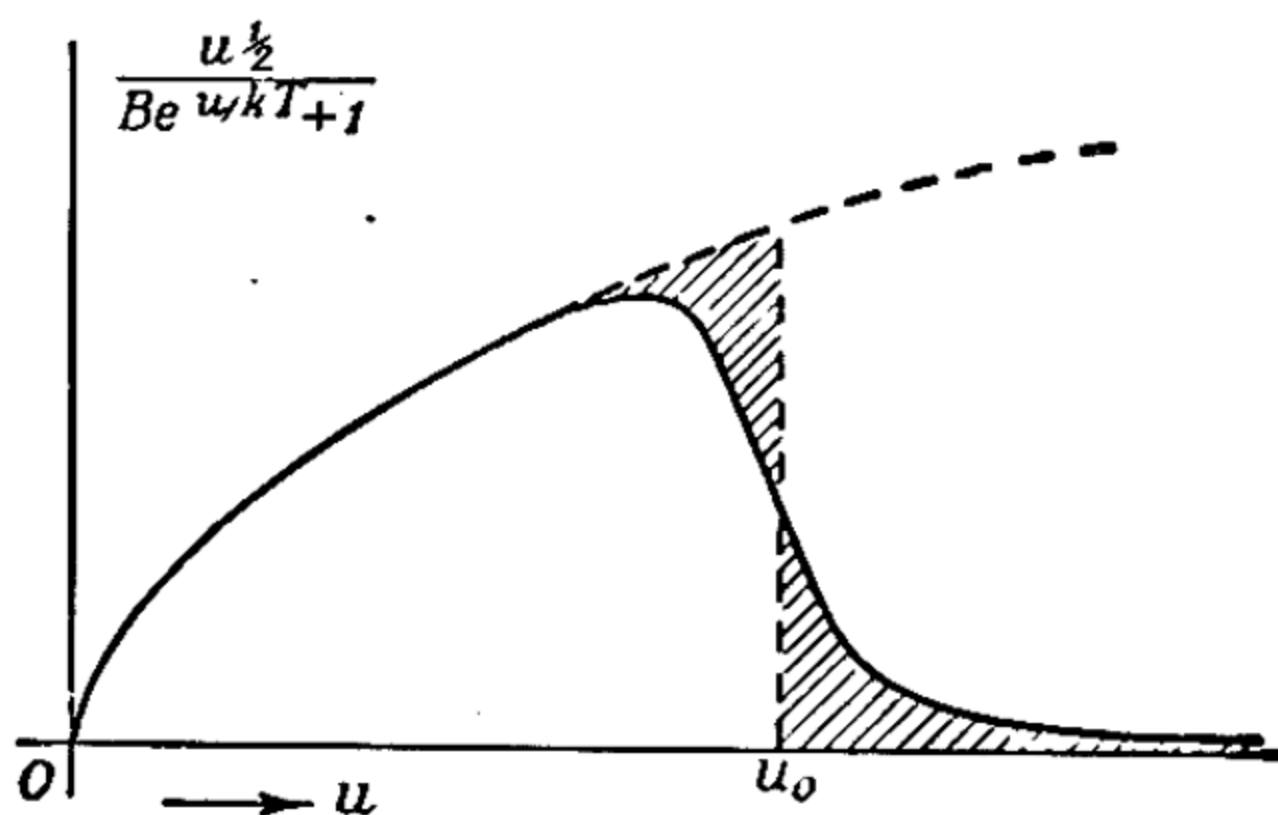


Fig. 1

equal in energy, i.e. the existence of the magnetic moment of the electron doubles the number of cells in each shell. We thus have

$$dZ_s = \frac{8\pi V \sqrt{2m^3 u}}{h^3} du_s \quad \dots \quad (28)$$

Inserting this value in equation (27) and omitting the subscript  $s$ , we obtain the number of electrons whose energy lies between  $u$  and  $u + du$ :

$$dN = \frac{8\pi V \sqrt{2m^3}}{h^3} \cdot \frac{u^{1/2}}{Be^{u/kT} + 1} du \quad \dots \quad (29)$$

We must now seek to express the constant  $B$  in terms of  $N$  by integrating over all values of the energy. The result is

$$N = \frac{8\pi V \sqrt{2m^3}}{h^3} \int_0^\infty \frac{u^{1/2} du}{Be^{u/kT} + 1} \quad \dots \quad (30)$$

Unfortunately the integral cannot be evaluated in closed form. We therefore avail ourselves of a somewhat crude approximation: if we

draw the curve of the integrand function it is seen to be a parabola as long as  $Be^{u/kT}$  can be neglected in comparison with 1 (cf. fig. 1), but as soon as the exponential becomes comparable with unity, the actual curve descends and finally approaches the axis of  $u$  asymptotically. If we erect a vertical line through the point  $u_0$ , which is determined by

$$Be^{u_0/kT} = 1 \quad \text{or} \quad u_0 = -kT \log B,$$

it is evident that the two shaded areas are approximately equal. We may therefore replace our integral approximately by the area under the parabola between 0 and  $u_0$ . In order to avoid difficulties with the sign, introduce the constant  $1/A$  in place of  $B$ , i.e.

$$-\log B = +\log A. \quad \cdot \cdot \cdot \cdot (31)$$

Then the area becomes

$$\int_0^\infty \frac{u^{1/2} du}{\frac{1}{A} e^{u/kT} + 1} \approx \frac{2}{3} (kT \log A)^{3/2} \quad \cdot \cdot (32)$$

Inserting this in (30) and solving for  $\log A$ , we obtain

$$\log A = \frac{h^2}{2mkT} \left( \frac{3N}{8\pi V} \right)^{2/3}, \quad \cdot \cdot \cdot \cdot (33)$$

and hence the distribution function becomes

$$dN = \frac{8\pi V \sqrt{2m^3}}{h^3} \cdot \frac{u^{1/2} du}{e^{u/kT - (h^2/2mkT)(3N/8\pi V)^{2/3}} + 1}. \quad \cdot (34)$$

The result becomes especially vivid if, instead of giving the number of electrons whose energy is between  $u$  and  $u + du$  and which are in the volume  $V$ , we specify their density in the momentum space. Since the shells have each a volume

$$d\tau = 4\pi p^2 dp = 4\pi \sqrt{2m^3 u} du,$$

the density of points in the momentum space becomes

$$\rho = \frac{dN}{d\tau} \propto \frac{1}{e^{u/kT - (h^2/2mkT)(3N/8\pi V)^{2/3}} + 1}. \quad \cdot \cdot \cdot (35)$$

For energy values less than

$$\frac{h^2}{2m} \left( \frac{3N}{8\pi V} \right)^{2/3}$$

the denominator has the value 1 at  $T = 0$ ; for energies greater than this value, it is infinite. Thus at  $T = 0$  the density of the representative points as a function of the corresponding energy is constant up to a certain point and then drops suddenly to zero. This result can be understood on the basis of the Pauli Principle. Since each cell may



contain but one electron, all energy levels must be occupied, from the lowest upward, until all electrons are provided for. Raising the temperature then has merely the effect of rounding the edges, as one may verify by inserting numerical values of  $T$  which are greater than zero.

In order to compute the energy of the electron gas we must evaluate the integral

$$U = \int_0^x u dN. \quad \cdot \cdot \cdot \cdot \cdot \cdot (36)$$

In the same way as before we obtain as a first approximation

$$\begin{aligned} U &= \frac{8\pi V \sqrt{2m^3}}{h^3} \int_0^\infty \frac{u^{3/2} du}{\frac{1}{A} e^{u/kT} + 1} \approx \frac{8\pi V \sqrt{2m^3}}{h^3} \int_0^{kT \log A} u^{3/2} du \\ &\approx \frac{8\pi V \sqrt{2m^3}}{h^3} \cdot \frac{2}{5} (kT \log A)^{5/2}. \quad \cdot \cdot \cdot \cdot \cdot \cdot (37) \end{aligned}$$

Inserting the value of  $\log A$ , we obtain, finally,

$$U \approx \frac{16\pi V \sqrt{2m^3} h^5 k^{5/2} T^{5/2}}{5h^3 (2m)^{5/2} k^{5/2} T^{5/2}} \cdot \left( \frac{3N}{8\pi V} \right)^{5/3} = \frac{4\pi V h^2}{5m} \left( \frac{3N}{8\pi V} \right)^{5/3}. \quad (38)$$

Thus to a first approximation the energy content of the electron gas is independent of the temperature; i.e. the energy found above is already present at the absolute zero. But since the specific heat is obtained by differentiating the energy with respect to  $T$ , it follows that the specific heat of the electron gas vanishes, to a first approximation. Thus the greatest obstacle in the electron theory of metallic conduction is eliminated, for the classical statistics leads to an increase of the specific heat of the solid metal from  $6R/2$  to  $9R/2$  per mol, assuming the same density of electrons as here. This increase is not observed. On the other hand, the Fermi-Dirac statistics leads to the observed absence of a contribution to the specific heat. It is a matter for a more complete calculation to compute the electrical and thermal conductivities of metals based on the new distribution. This has been done by Sommerfeld and his co-workers in a series of papers, but the details cannot be given here.\* A further advance was made by taking into account the interaction with the lattice ions by the introduction of a spatially periodic potential. This computation must be based on quantum mechanics and so it is described later (Chap. XLI, p. 707 *et seq.*).

In this way we arrive at a semiquantitative understanding of *all* electrical and optical properties of metals except one, whose explanation is lacking, viz. *superconductivity*. This consists in the fact that at a few degrees above absolute zero the metals of columns IIb, IIIb, IVa, IVb, Va and VIa of the periodic table (p. 676) as well as several alloys suddenly lose all electrical resistance. The transition temperature is a characteristic of the material. Connected with the disappearance of resistivity is the disappearance of the magnetic permeability.

\* Consult A. Sommerfeld, *Zeitschr. f. Phys.*, 47 1-2, pp. 1, 43 (1928).

#### 4. Entropy Constant and Chemical Constant of an Ideal Monatomic Gas.

The thermodynamics of the vapour pressure curve and of chemical equilibrium peremptorily demands that the absolute value of the entropy should be determined (cf. p. 553). Since the Quantum Theory leads to a determination of the size of the element of phase space, it is to be expected that the absolute value of the entropy may be calculated readily, for the size of the element enters into this quantity (cf. p. 563). However, a difficulty exists in respect that the classical quantum statistics yields a quantum condition only for periodic motions. Now the translational motion of the atoms of an ideal gas (we restrict the present discussion to ideal monatomic gases) is entirely non-periodic; if we wish to apply the Quantum Theory we must artificially introduce a periodicity, say by treating the to and fro motion between the walls as a periodic motion. Moreover, in order to arrive at a reasonable result with such considerations, we find ourselves facing an inconsistency with the Boltzmann statistics; here we do not differentiate between two states resulting from the interchange of two atoms, while in Boltzmann's method two such complexions are taken to constitute two different micro-states. For this reason we refrained from deriving the chemical constant on the basis of the classical quantum statistics (Boltzmann statistics augmented by the quantum condition), since this actually uses the principal feature of both the Bose and Fermi methods, viz. the absence of any distinction between like atoms, which obtains even in the micro-state. The absolute value of the entropy follows in a simple manner from the postulates of the new statistics. The Bose and the Fermi statistics both lead to the same result, since according to equation (23) (p. 631), the constant  $B$  for ordinary gases is large compared with unity, and so the "one" in the denominator of each distribution function may be dropped. This makes the two forms identical.

In order to derive the entropy constant we start from the formulæ of § 2 (p. 629):

$$S = k \sum_s Z_s \log Z_s - k \sum_s \sum_n Z_{ns} \log Z_{ns}, \quad . . . \quad (5)$$

with the auxiliary conditions

$$\sum_n Z_{ns} = Z_s, \quad . . . . . \quad (6)$$

$$\sum_s \sum_n n Z_{ns} = N = \sum_s N_s, \quad . . . . . \quad (7)$$

$$\sum_s \sum_n n u_s Z_{ns} = U = \sum_s N_s u_s. \quad . . . . . \quad (8)$$

There we found the most probable distribution to be

$$Z_{ns} = \alpha_s e^{-n(\beta + u_s/kT)}, \quad \dots \quad (13)$$

$$Z_s = \alpha_s (1 - e^{-\beta - u_s/kT})^{-1}, \quad \dots \quad (14)$$

$$N_s = Z_s (e^{\beta + u_s/kT} - 1)^{-1}. \quad \dots \quad (16)$$

Under the assumption that  $B = e^\beta \gg 1$ , i.e.  $1/B = e^{-\beta} \ll 1$ , we obtain

$$\log Z_s = \log \alpha_s + e^{-\beta - u_s/kT} = \log \alpha_s + \frac{1}{B} e^{-u_s/kT}, \quad \dots \quad (39)$$

$$N_s = \frac{Z_s}{B} e^{-u_s/kT}. \quad \dots \quad (40)$$

If we insert the values from (39) and (13) in the logarithms of equation (5), we obtain

$$\begin{aligned} S = k \sum_s Z_s \log \alpha_s + k \sum_s \frac{Z_s}{B} e^{-u_s/kT} - k \sum_s \sum_n Z_{ns} \log \alpha_s + k \sum_s \sum_n Z_{ns} n \beta \\ + k \sum_s \sum_n Z_{ns} n \frac{u_s}{kT}. \quad \dots \quad (41) \end{aligned}$$

According to equation (6), the first and third terms drop out. By equation (40), the second term is equivalent to  $k \sum N_s = kN$ ; by (7), the fourth term is  $k\beta N$ , and according to (8) the last term is  $U/T$ . Thus the equation simplifies to

$$S = kN + k\beta N + \frac{U}{T}. \quad \dots \quad (42)$$

Taking one mol, i.e.  $L$  atoms, and putting  $u = 3RT/2$ , we obtain

$$s = R + \beta R + \frac{3}{2}R = \beta R + \frac{5}{2}R. \quad \dots \quad (43)$$

Using the expression for  $\beta = \log B$  given by equation (23) (p. 631) we obtain  $s$  as a function of  $v$  and  $T$ :

$$\frac{s}{R} = \log v + \frac{3}{2} \log T + \frac{5}{2} + \log \frac{(2\pi mk)^{3/2}}{Lh^3}. \quad \dots \quad (44)$$

In the definition of the chemical constant,  $s$  is assumed to be a function of  $p$  and  $T$ . Performing the conversion with the help of the equation of state we have, since  $R = kL$ ,

$$\frac{s}{R} = \frac{5}{2} \log T - \log p + \frac{5}{2} + \log \frac{(2\pi m)^{3/2} k^{5/2}}{h^3} = \frac{5}{2} \log T - \log p + \frac{s_0}{R}, \quad (45)$$

and since  $c_p = 5R/2$ , the chemical constant of an ideal monatomic gas becomes

$$i_{\text{abs}} = \frac{s_0 - c_p}{R} = \log \frac{(2\pi m)^{3/2} k^{5/2}}{h^3}. \quad \dots \quad (46)$$



Inserting the numerical values of the atomic quantities and introducing the logarithms of the factors needed for the conversion to practical units ( $p$  in atm.), we obtain the formula quoted on p. 538:

$$i = -1.59 + \frac{3}{2} \log A. \quad . \quad . \quad . \quad . \quad (47)$$

*Ex. 131. Thermal Ionization:* Calculate the chemical constant of the electron gas in the non-degenerate (rarefied) state and use the result to compute the thermal equilibrium between atoms, ions and electrons in a hot gas in terms of the work of ionization of the atom.

## **PART VII**

### **THE STRUCTURE OF ATOMS AND MOLECULES AND THE THEORY OF SPECTRA**





## CHAPTER XXXIX

### THE MECHANICS OF SIMPLE ATOMIC MODELS

#### 1. Investigation of the Structure of the Atom; the Scattering of $\alpha$ -Particles.

The many phenomena in which electrons are released from neutral atoms or atomic aggregates—e.g. the photoelectric effect and the phenomena of electrical discharge in gases—suggested from the very first that the atom is a structure made up of electrons, the charge of which is compensated by an equal number of positive elementary charges. The question immediately arises of how the two kinds of charge are arranged within the atom, and where the inertial mass is located. This question was attacked experimentally by probing or sounding the atom by means of projectiles of subatomic size. These were sent through the atoms, and the arrangement of the charges was deduced from the alterations in the paths of the particles. Lenard, who first undertook experiments of this kind, used electrons as projectiles. He recognized the fact that the volume occupied by an atom in gas-kinetic collisions or in liquids and solids is merely the effect of electrical fields of force due to charges within the atom, while the intrinsic volume of the parts of the atom is vanishingly small. Rutherford and his associates obtained particularly simple conditions by using  $\alpha$ -particles expelled from radioactive substances instead of electrons. The former carry a positive charge of two elementary units and, like the electrons, are of exceedingly small size, but have a mass almost 7400 times as great—the same as that of a helium atom. For this reason, the deflection of the path of an  $\alpha$ -particle requires a centre having a large inertial mass, inasmuch as the conservation of linear momentum requires that the centre of gravity of the two bodies should continue in its original motion. For a collision between an  $\alpha$ -particle and an electron, the centre of gravity is practically coincident with the former, so that an electron is not capable of appreciably deflecting an  $\alpha$ -particle. Hence the use of  $\alpha$ -particles in this way is able to furnish information concerning the location of the mass of the atom, so far as it is connected with electric charges. C. T. R. Wilson found \* that if ions are present in supersaturated dust-free air they

\* C. T. R. Wilson, *Phil. Trans.*, A, 189, p. 265 (1897).

act as centres of condensation for water droplets. This is the principle of Wilson's so-called "cloud chamber". If the ions are produced by the passage of  $\alpha$ -particles through the air, the path of the particle is outlined as a "fog track". Photographs of such tracks show that a deflection through a large angle occurs suddenly, rather than by the cumulative effect of many smaller deflections. This further simplifies the theory of the effect to the extent that we may limit our calculations to simple scattering and omit from consideration the rare case of a single particle experiencing multiple deflections.

Consider now what happens to a single particle. We shall ascribe to the scattering centre a mass so great that we may consider it at rest during the scattering process. Since the central mass can in no instance be greater than that of the atom, this holds only for scattering by atoms of heavy elements. Actually, the first experiments were made with such elements. Assume the scattering centre to have  $Z$  elementary charges. Then the force acting on the  $\alpha$ -particle is  $2Ze^2/r^2$ . Since this law of force is the same as that of gravitation, so far as the dependence on  $r$  is concerned, we can apply the results of p. 91 at once. The path is thus a conic section with the scattering centre at one focus. This leaves the question of the sign of the central charge open. The hyperbolic orbits in either case differ only to the extent that the central body occupies the focus on the concave side when the force is attractive, and is at the focal point on the convex side when the force is one of repulsion. In harmony with other experiments, in which the positive charge is always associated with masses which are integral multiples of that of a hydrogen atom, we assume the central charge to be positive. Comparing the potential energy of gravitation  $-\gamma mM/r$  with the electrostatic energy  $+2Ze^2/r$ , we see that  $\gamma M$  must be replaced by  $-2Ze^2/m$  in the formulæ of p. 90, where  $m$  is the mass of the moving  $\alpha$ -particle. In order to avoid confusing it with the mass of the electron, we give it the subscript  $\alpha$  in what follows. According to equation (21) (p. 91), the equation of the orbit becomes

$$r = \frac{-\frac{2c^2m_\alpha}{e^2Z}}{1 + \frac{cm_\alpha}{Ze^2}\sqrt{v_0^2 + \frac{4Ze^2}{m_\alpha r_0} + \frac{Z^2e^4}{m_\alpha^2c^2}\cos\phi}}, \quad \dots (1)$$

where  $c$  is the areal velocity. The negative sign signifies that if the positive value of the radical is taken, the branch of the hyperbola directed away from the centre is the one described by the particle. If, for example, we take  $\phi = 0$ , the corresponding value of  $r$  in fig. 1 is not to be drawn to the right, but to the left. At the beginning of the motion the particle is at an infinite distance, i.e.  $1/r$  is zero, and its velocity is  $v_0$ . If the initial direction, which is represented by the

asymptote of the hyperbola, is a distance  $p$  from the centre, the areal velocity at the beginning is  $v_0 p/2$ , since the curve and the asymptote are practically coincident at great distances from the centre. This areal velocity is represented by the shaded triangle in fig. 1. Because

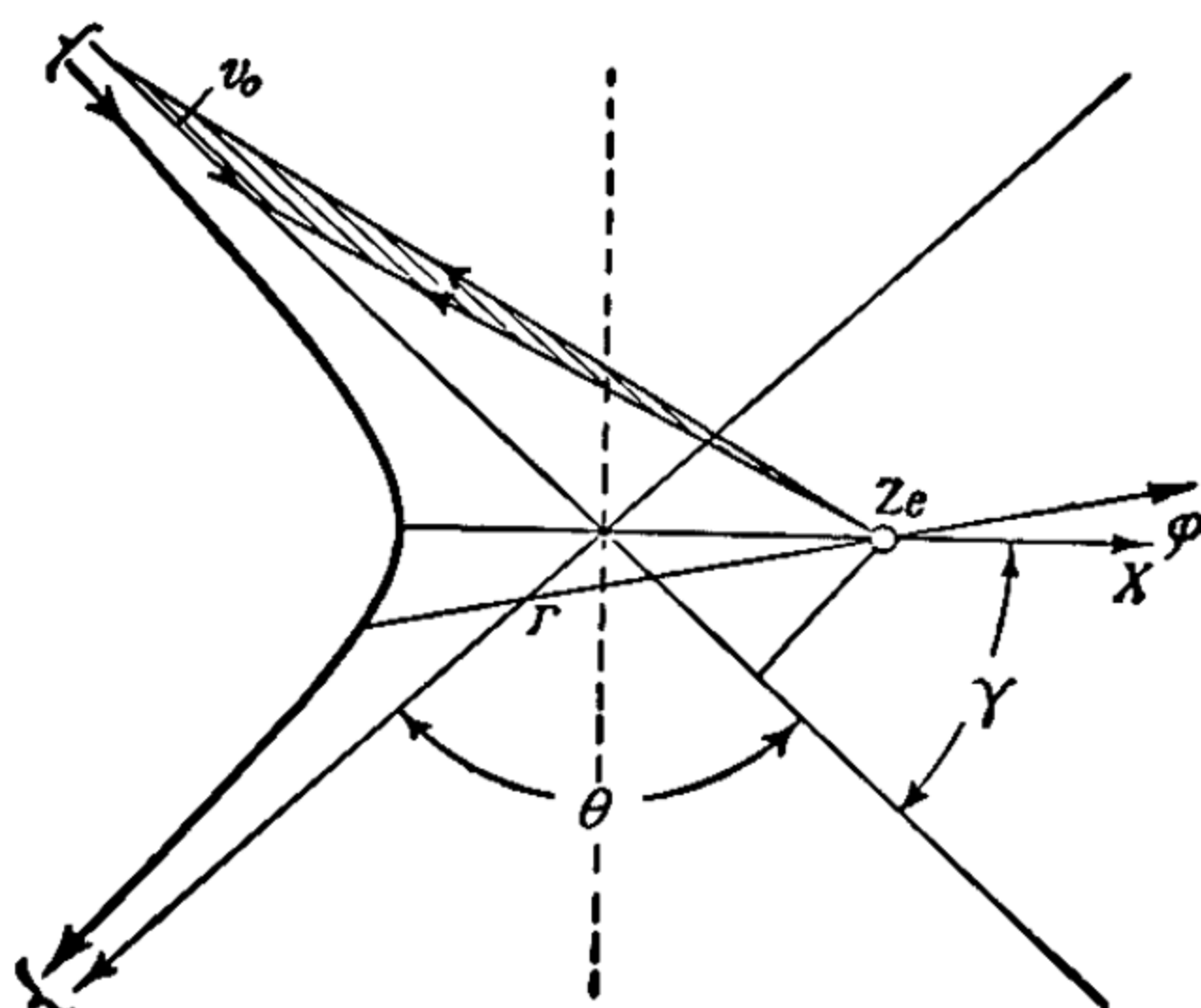


Fig. 1

of the constancy of the areal velocity we may put this value in equation (1) in place of  $c$ , thus obtaining

$$r = \frac{-\frac{v_0^2 p^2 m_a}{2e^2 Z}}{1 + \frac{v_0 p m_a}{2Ze^2} \sqrt{v_0^2 + \frac{4Z^2 e^4}{m_a^2 v_0^2 p^2} \cos \phi}} \quad \dots \quad (2)$$

The numerical eccentricity becomes

$$\epsilon = \frac{v_0 p m_a}{2Ze^2} \sqrt{v_0^2 + \frac{4Z^2 e^4}{v_0^2 p^2 m_a^2}} \quad \dots \quad (3)$$

Denote the acute angle between the initial direction of the particle and the positive  $x$ -axis by  $\gamma$ . Then the angle of deviation, i.e. the angle between the asymptote and the one corresponding to the final direction of motion, is given by

$$\frac{\theta}{2} = \frac{\pi}{2} - \gamma. \quad \dots \quad (4)$$

From the geometry of the hyperbola,\*

$$\cos(\pi - \gamma) = -\frac{1}{\epsilon}, \quad \cot \gamma = \tan \frac{\theta}{2} = \frac{1}{\sqrt{\epsilon^2 - 1}},$$

\* It follows from the polar equation  $r = -p/(1 + \epsilon \cos \phi)$  that  $\cos \phi = -1$ , when  $r = \infty$ .



so that half the angle of deviation is given by

$$\tan \frac{\theta}{2} = \frac{1}{\sqrt{\frac{v_0^2 p^2 m_a^2}{4Z^2 e^4} \left( v_0^2 + \frac{4Z^2 e^4}{v_0^2 p^2 m_a^2} \right) - 1}} = \frac{2Ze^2}{v_0^2 m_a p}. \quad (5)$$

This simple formula gives the angle of deviation as a function of the parameter  $p$ ; it is impossible, however, to test the validity of the equation experimentally, since we cannot directly determine the distance of the initial direction of a given particle from an atomic nucleus. The result of the computation can be tested only by statistical methods, in which the theory of probability is applied to the distribution of the various parameters in order to determine that of the angle of deflection. In fact, this marks a weakness in logic of the entire atomic mechanics, as based on models: the quantities necessary to determine the course of a single process cannot be determined exactly;\* then, if we pass to a statistical distribution, the quantity  $p$ , which cannot be observed, again drops out of the calculations. Thus the elementary process is used only as an intermediate concept. The direct procedure would be to assume that a pencil of  $\alpha$ -rays passes into the field of atomic nuclei and to inquire into the distribution of the rays after passing through this region. This is, in fact, the way in which such problems are attacked by the newer quantum mechanics to be discussed in the following chapter; there an attempt is made to operate only with directly observable quantities. For the moment, however, we shall continue to discuss the atomic models, and shall now compute the angular distribution of scattered particles by means of such concepts.

If the layer of atoms which is penetrated has a thickness  $s$ , and if there are  $N$  atomic centres per unit volume, then a hollow cylinder of radii  $p$  and  $p + \Delta p$  whose axis is in the direction of the incident rays contains, on the average,

$$\Delta N = 2\pi p s N \Delta p \quad . . . . . (6)$$

nuclei. For the small values of  $p$  which are considered, this is a very small number, e.g. 0.001. The introduction of an average value is justified by the fact that the primary pencil has a finite cross-section, so that the layer is penetrated at various points, which is equivalent to using an infinitely narrow pencil of rays and moving the bombarded layer of atoms to and fro within the same limits of variability. Assume that each particle is scattered once; then, of  $n$  primary particles, a number

$$\Delta n = n . 2\pi p s N \Delta p \quad . . . . . (7)$$

\* Later we shall find that these quantities are inherently indeterminable, quite apart from the insufficiency of our methods of attack. Cf. p. 690.

will suffer "impacts" with a nucleus lying between  $p$  and  $p + \Delta p$ . Now by (5),  $p$  is connected with the angle of scattering,  $\theta$ , by the formula

$$\left. \begin{aligned} p &= \frac{2e^2Z}{v_0^2 m_\alpha} \cot \frac{\theta}{2} \\ \Delta p &= \frac{e^2Z \Delta \theta}{v_0^2 m_\alpha \sin^2 \theta/2} \end{aligned} \right\}, \quad \cdot \cdot \cdot \cdot \cdot \quad (8)$$

so that the fractional number  $\Delta n/n$  of  $\alpha$ -particles scattered between the angles  $\theta$  and  $\theta + \Delta \theta$  is

$$\frac{\Delta n}{n} = \frac{4\pi s N e^4 Z^2 \cos \frac{\theta}{2}}{v_0^4 m_\alpha^2 \sin^3 \frac{\theta}{2}} \Delta \theta. \quad \cdot \cdot \cdot \cdot \cdot \quad (9)$$

Formula (9) becomes invalid when  $\theta = 0$ , for then it yields  $\Delta n/n = \infty$ . This would mean that practically all  $\alpha$ -particles go through without being scattered. The discrepancy is explained by the fact that we are not permitted to take an infinitely large upper limit for  $p$  but must use the finite value given by the mean distance apart of the atoms of the scattering material.

## 2. Conclusions from the Experiments on the Scattering of $\alpha$ -Particles.

The scattering formula (9) is well confirmed for the heavier elements as scattering centres and for *finite* angles of deflection. The experimental verification of the scattering formula justifies, for one thing, our hypothesis of electrostatic forces between the particles and the nuclei and confirms the assumption of simple scattering. Further, a comparison of the scattering powers of various substances shows that  $N$  may be set equal to the number of atoms per unit volume and that  $Z$  coincides with the place of the element in the periodic system. We shall discuss this important result in greater detail below.

Deflections of  $90^\circ$  are sometimes observed. If we use equation (9) to determine the value of  $p$  corresponding to  $\alpha$ -particle velocities of  $10^9$  cm./sec. and to a deflection angle of  $90^\circ$ , we obtain a value of the order of  $10^{-12}$  cm. This defines the radius of the sphere within which the inertial mass of the atom is concentrated, and the Coulomb law of force is still valid at these distances. We shall refer to this central structure as the *nucleus* of the atom. Its size is vanishingly small compared with the gas-kinetic radius of the atom, which is of the order of magnitude of  $10^{-8}$  cm. This extension in space must therefore



be maintained by the electrons surrounding the nucleus; their electrical fields of force prevent the closer approach of other atoms.

However, the most important result of the scattering experiments is the recognition of the fact that *the nuclear charge  $Z$  consists of a number of positive elementary charges equal to the position number of the element in the periodic table*. Formerly, the elements were arranged in the periodic system in the order of atomic weight. At certain places this sequence had to be violated in order to satisfy the sequential arrangement according to certain chemical properties. We now recognize that the nuclear charge is the fundamental parameter, and this makes the order correct at the places mentioned. Since the atom as a whole is normally electrically neutral, the number of electrons outside the nucleus must also be equal to  $Z$ . The position number  $Z$  of the element was called the *atomic number* by Moseley (cf. p. 655). We thus have the important relation:

$$\begin{aligned}\text{Nuclear charge} &= \text{Number of electrons outside the nucleus} \\ &= \text{Atomic number of the element.}\end{aligned}$$

Hydrogen, of atomic number 1, has a nucleus with *one* elementary charge whose mass is that of a hydrogen atom (less the mass of the electron).  $M_H$  is  $1.66 \times 10^{-24}$  gm. The hydrogen nucleus is to be looked upon as the atom of positive electricity, which differs from its counterpart—the electron—in having a much larger mass. The nucleus of the hydrogen atom is called a *proton*. Further, since helium has the atomic number 2, and since the  $\alpha$ -particle has a resultant charge of  $+2$ , and a mass equal to that of the helium atom, we recognize the  $\alpha$ -particle to be a *helium nucleus*. One might conclude that the nuclei of the heavier elements consist of protons only, but there are two serious objections to this idea: since the mass of each proton is that of a hydrogen atom (the mass of the electron may be neglected in comparison), the atomic number would have to be equal to the atomic weight, which is obviously not the case. Prior to the discovery of the neutron (p. 750) it was assumed that electrons were present in the nucleus itself, thus serving to reduce its net charge. Supporting this view was the fact that in many radioactive changes—which are certainly nuclear processes—electrons are emitted in the form of  $\beta$ -rays. However, the assumption that electrons exist in the nucleus led to such great difficulties that it had to be abandoned, and now we explain the difference between atomic weight and nuclear charge in terms of neutrons present in the nucleus.

The second objection—that if the elements are built up of protons and electrons all atomic weights, referred to hydrogen, must be whole numbers (which is evidently not so)—was removed by the discovery of isotopes. As shown especially by Aston in his analysis of canal rays



with the mass spectrograph (p. 441), the chemical elements are largely mixtures of atoms of different atomic weights, but having the same nuclear charge and hence the same number of outer electrons. Apart from small deviations, the several atomic weights are integral multiples not of that of hydrogen itself, but of a weight about 8 per cent smaller, i.e.  $1/16$  the atomic weight of oxygen. The small departures from the integral multiples are to be interpreted on the basis of the principle of the inertia of energy as differences in the energy of formation of the nuclei. The large value of these energy changes (obtained by multiplying the difference in mass by the square of the velocity of light) corresponds with the order of magnitude of the energy released in radioactive processes.

### 3. The Bohr Model of the Hydrogen Atom.

Now that we have before us a general rough picture of the atom in the form of a heavy nucleus about which electrons cluster, we naturally desire more definite information concerning the arrangement of these electrons. The first question which arises pertains to the nature of the force which balances the electrostatic attraction of the nucleus, thus preventing the electrons from falling into it. Similar situations in celestial mechanics suggest that the opposing force is centrifugal force. We are thus led to assume that the electrons execute rapid motions of revolution about the nucleus. This idea was followed out mathematically by Bohr,\* first for the simplest case of the hydrogen atom, and with the simplest form of electron orbit—the circle. There is one great difficulty. The revolving electron, together with the positive nucleus, represents a miniature antenna in which two vibrations which differ in phase by  $\pi/2$  are taking place. According to the laws of electrodynamics, an oscillator of this kind must radiate energy in the form of electromagnetic waves. Such an energy loss can occur only at the expense of the potential energy, i.e. the orbits must decrease in size, so that the electrons traverse spiral orbits which ultimately lead into the nucleus. This cataclysm must take place within time intervals of the order of  $10^{-8}$  sec., so that stable atoms of this kind are out of the question. But the validity of the electrodynamic laws has been demonstrated experimentally only for macroscopic antennæ; it may be that other laws obtain for the individually unobservable atomic radiation processes, which are limited only by the requirement that the average effect be in harmony with the macroscopic electrodynamic results. Bohr thus made a first postulate to the effect that of the continuous sequence of mechanically possible orbits, only a discrete set is capable of existence. These orbits are assumed

\* N. Bohr, *Phil. Mag.*, 26, p. 1 (1913).

to have the property that no radiation takes place when the electron is in one of them. The criterion for such an orbit is taken to be the condition that the phase integral (p. 130) is an integral multiple of Planck's quantum of action:

$$J = \oint p dq = nh. \quad . . . . . (10)$$

The second postulate is that the emission or absorption of radiation (light) takes place according as the atom passes from a higher to a lower energy state or vice versa. At the same time a quantitative relation found experimentally to hold for all elementary energy transformations, e.g. the photoelectric effect, must be satisfied: the energy difference of the two states must be equal to the quantum  $h\nu$  of the emitted or absorbed radiation. No statement is made regarding the actual mechanism of the process of radiation or absorption. Also, there is apparently no connexion between the frequency of the radiation and the frequency of revolution of the electron in its orbit; we shall find, however, in § 6 (p. 655) that there is an asymptotic connexion between the two.

Let us now apply the two postulates to the hydrogen atom, whose electron revolves about the nucleus in a circular orbit of radius  $a$ . The kinetic energy  $T$  amounts to

$$T = \frac{1}{2} ma^2 \omega^2 = \frac{1}{2} ma^2 \dot{\phi}^2, \quad . . . . . (11)$$

so that the momentum co-ordinate is

$$p_\phi = ma^2 \dot{\phi} = ma^2 \omega. \quad . . . . . (12)$$

The phase integral is to be extended over an entire circuit of the position co-ordinate—in this case from 0 to  $2\pi$ :

$$J = \oint p dq = \int_0^{2\pi} ma^2 \omega d\phi = 2\pi ma^2 \omega. \quad . . (13)$$

We then have the quantum condition

$$2\pi ma^2 \omega = nh. \quad . . . . . (14)$$

In addition, there is the equation of mechanical equilibrium

$$\frac{e^2}{a^2} = ma\omega^2. \quad . . . . . (15)$$

There are thus two equations for the two unknowns  $a$  and  $\omega$ . Solving for these quantities, we obtain

$$a = \frac{n^2 h^2}{4\pi^2 m e^2}. \quad . . . . . (16)$$

and

$$\omega = \frac{8\pi^3 m e^4}{n^3 h^3}. \quad . . . . . (17)$$

C

Inserting numerical values for  $e$ ,  $h$  and  $m$ , we obtain for  $n = 1$  the value

$$a_1 = 0.528 \times 10^{-8} \text{ cm.}, \quad . \quad . \quad . \quad . \quad . \quad (18)$$

i.e. a value which corresponds very well with the gas-kinetic atomic radii. The fact that a purely chance agreement between these quantities is highly improbable lends plausibility to this model of the atom. The values of  $\omega$  correspond to the frequencies of ultra-violet light; although the radiated frequencies have no immediate connexion with  $\omega$ , it seems significant that they are of the same order of magnitude. In order to compute the actual radiated frequencies on the basis of the second Bohr postulate, we must write down the energy in the  $n$ th quantum state:

$$\begin{aligned} E_n = T + U &= \frac{m}{2} a_n^2 \omega_n^2 - \frac{e^2}{a_n} = \frac{2\pi^2 m e^4}{n^2 h^2} - \frac{4\pi^2 m e^4}{n^2 h^2} \\ &= - \frac{2\pi^2 m e^4}{n^2 h^2}. \quad . \quad . \quad . \quad . \quad . \quad (19) \end{aligned}$$

The negative sign arises as a result of taking the zero point of the potential energy at infinity. The potential energy, which is exactly twice the kinetic energy, is negative in the finite region, since work is obtained when the distance between the nucleus and the electron decreases from infinity. The *magnitude* of the energy with a positive sign signifies the work which must be done in order to remove the electron from its orbit to infinity, i.e. to detach it from the atom. This work is usually expressed in "electron volts", i.e. the difference of potential (in volts) through which an electron must fall in order to obtain this amount of kinetic energy. The *ionization potential* of an atom is the name given to the energy difference, expressed in electron volts, between the normal state and the state of complete removal. The former is the state of lowest energy—the one having the lowest quantum number; for hydrogen this is 1. The *excitation potential* of a given state is the difference in energy between this and the "ground" state.

According to the second Bohr postulate the frequency of the light emitted or absorbed in a transition from one state of the atom to another is given by

$$\nu = \frac{1}{h} (E_i - E_f) = \frac{2\pi^2 e^4 m}{h^3} \left( \frac{1}{n_f^2} - \frac{1}{n_i^2} \right), \quad . \quad . \quad (20)$$

where the subscripts  $i$  and  $f$  denote the initial and final states, respectively, for the case of emission. In point of fact, all the spectral lines emitted by the hydrogen atom do actually satisfy this law. The sequence of lines with  $n_f = 2$ ,  $n_i = 3, 4, 5, \dots$ , four lines of which are in the visible region, has been known for a long time as Balmer's



series. The lines crowd closer together as the "current number"  $n_i$  increases, and at the series limit—given by  $n_i = \infty$ , i.e. vanishing variable term—there is an accumulation point of spectral lines, in the mathematical sense. Actually the intensity of the lines decreases as the series limit is approached. The

series of lines belonging to  $n_f = 1$ ,  $n_i = 2, 3, 4, \dots$  lies in the far ultra-violet region, and was not found until later by Lyman. The lines corresponding to  $n_f = 3$ ,  $n_i = 4, 5, \dots$  and to  $n_f = 4$ ,  $n_i = 5, 6, 7, \dots$  are in the infra-red, and are named after their discoverers the Paschen Series and the Brackett Series respectively.

The origin of the individual lines is best visualized with the aid of an "energy level diagram" as in fig. 2, in which the various energy levels which can be assumed by the atom according to the first Bohr postulate are represented by properly spaced horizontal levels. The transitions are represented by arrows in the direction of transition. The central numbers are wave-lengths in Ångströms.

An outstanding success of the theory is to be seen in the fact that it not only yields the *form* of the law obeyed by the hydrogen lines, but also gives the correct numerical value of the constant of the formula, when the numerical values of  $e$ ,  $m$  and  $h$  are inserted. It is true that the accuracy to which  $e$ ,  $m$  and  $h$  are known is far below

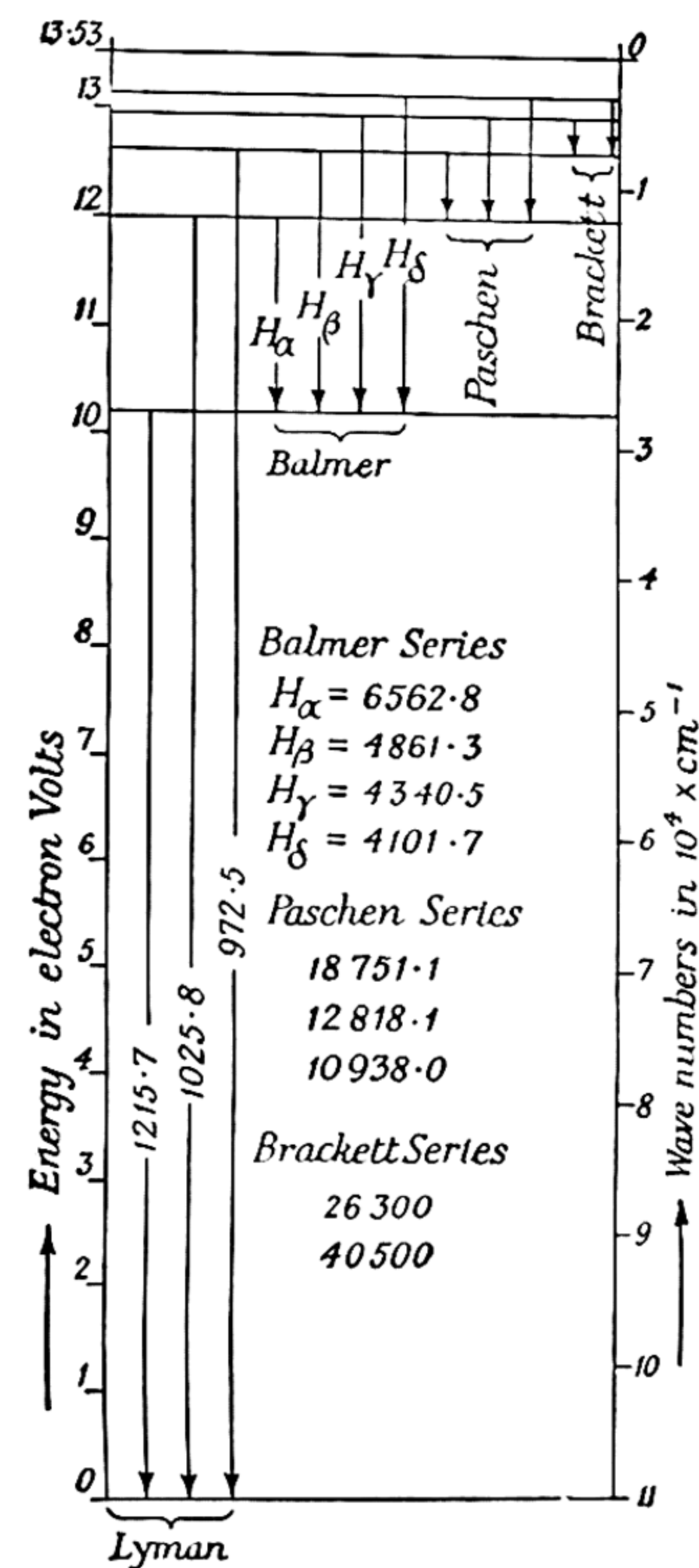


Fig. 2

that available in spectroscopic measurements, so that once we have attained confidence in the theory, we shall rather use the constant to determine  $e$ ,  $m$  and  $h$ .

It is customary in spectroscopy to use the reciprocal of the wave-length,  $1/\lambda$ , rather than the actual frequency,  $c/\lambda$ . These so-called *wave numbers* are also designated by  $\nu$ . In order to avoid confusion, we write  $\nu$  with a bar over it to denote wave numbers. If the Balmer formula is written in these units,

$$\bar{\nu} = R \left( \frac{1}{n_f^2} - \frac{1}{n_i^2} \right). \quad \dots \dots \dots (21)$$

the constant  $R$  is called the *Rydberg Number*. By equation (20), this constant is given by

$$R = \frac{2\pi^2 e^4 m}{ch^3} \text{ cm.}^{-1}, \quad \dots \dots \dots (22)$$

and its numerical value is

$$R = 109737 \text{ cm.}^{-1}.$$

#### 4. Consideration of the Motion of the Nucleus. The Spectrum of Hydrogen and the Spectrum of Ionized Helium.

Even if we neglect the relativistic variability of the electron mass, the simple formulæ derived in § 3 (p. 649) are in need of refinement, since we assumed the nucleus to be at rest. This is true only for infinitely heavy nuclei; in reality, the motion takes place about the common centre of gravity of nucleus and electron, and in computing the energy we must take into account the kinetic energy of the motion of the nucleus also. Denoting the mass of the nucleus by  $M$ , its distance from the centre of gravity by  $A$ , and the distance of the electron from this point by  $a$ , we have

$$MA = ma. \quad \dots \dots \dots (23)$$

The condition of mechanical equilibrium for the nucleus is

$$\frac{e^2}{(a + A)^2} = MA\omega^2, \quad \dots \dots \dots (24)$$

and for the electron,

$$\frac{e^2}{(a + A)^2} = ma\omega^2. \quad \dots \dots \dots (25)$$

It is found expedient to express both  $a$  and  $A$  in terms of the distance  $(a + A)$  between nucleus and electron. From equation (23) it follows that

$$a + A = a \left( 1 + \frac{m}{M} \right), \quad \text{i.e.} \quad a = \frac{a + A}{1 + \frac{m}{M}} \quad \text{and} \quad A = \frac{a + A}{1 + \frac{m}{M}} \cdot \frac{m}{M}. \quad (26)$$

We define a "reduced" mass  $\mu$  by means of the equation

$$\frac{1}{\mu} = \frac{1}{m} + \frac{1}{M}. \quad \dots \dots \dots (27)$$

Then

$$1 + \frac{m}{M} = \frac{m}{\mu}, \quad \text{i.e.} \quad a = (a + A) \frac{\mu}{m}. \quad \dots \dots \dots (28)$$

and the condition of equilibrium for the electron becomes

$$\frac{e^2}{(a + A)^2} = \mu(a + A)\omega^2. \quad \cdot \cdot \cdot \cdot (29)$$

The total energy becomes

$$T + U = \frac{\omega^2}{2} (MA^2 + ma^2) - \frac{e^2}{a + A} = \frac{\mu\omega^2}{2} (a + A)^2 - \frac{e^2}{a + A}. \quad (30)$$

Comparing these formulæ with the corresponding ones of § 3, we see that they become identical if we replace the electron mass  $m$  by the reduced mass  $\mu$  and if we understand  $a_n$  in § 3 to mean the distance between the nucleus and the electron. Then the Rydberg constant for hydrogen becomes

$$R_H = \frac{2\pi^2 e^4 \mu_H}{ch^3} = 109677.69 \text{ cm.}^{-1}, \quad \left( \frac{1}{\mu_H} = \frac{1}{m} + \frac{1}{M_H} \right). \quad (31)$$

For an element with infinite nuclear mass we have the former value of this constant:

$$R = R_\infty = \frac{2\pi^2 e^4 m}{ch^3}. \quad \cdot \cdot \cdot \cdot (22)$$

For a one-electron system whose nucleus has the mass of the helium nucleus, we have

$$R_{He} = \frac{2\pi^2 e^4 \mu_{He}}{ch^3}, \quad \left( \frac{1}{\mu_{He}} = \frac{1}{m} + \frac{1}{M_{He}} \right). \quad \cdot \cdot (31')$$

The spectrum of this system is found in nature. If one of the two electrons of the helium atom is removed by violent excitation we have such a system; electrically this structure differs from a hydrogen atom only in having twice the nuclear charge. This means that the factor 2 must be supplied to  $e^2$  in the equation of mechanical equilibrium. In general, for a nucleus having a charge  $Z$ , the factor  $Z$  must be inserted. The formula for the energy levels contains  $e^4$ , so that the factor 4 (in general,  $Z^2$ ) must be supplied; the energy levels of the singly-ionized helium atom ( $He^+$ ) become:\*

$$E_n = \frac{4 \times 2\pi^2 e^4 \mu_{He}}{n^2 h^2} = \frac{4R_{He}hc}{n^2}. \quad \cdot \cdot \cdot (32)$$

Thus the wave-lengths of the lines which correspond to the hydrogen lines are one-fourth those of the hydrogen lines, so that the  $He^+$  spectrum corresponding to the Lyman and Balmer series lies in the far ultra-violet. For  $n_f = 4$ , however, the lines fall in the visible region.

\* The negative sign is omitted before  $E_n$ , since it is customary in spectroscopy to reckon the term values as representing work done in removing an electron.



Writing this series in the form

$$\nu = 4R_{He} \left( \frac{1}{4^2} - \frac{1}{n^2} \right) = R_{He} \left\{ \frac{1}{2^2} - \frac{1}{\left(\frac{n}{2}\right)^2} \right\}, \quad \dots \quad (33)$$

we recognize that alternate lines must coincide approximately with hydrogen lines. This is the so-called Pickering series, which has been known for a long time in stellar spectroscopy. It was formerly attributed to hydrogen, inasmuch as so close a connexion between the spectra of two different elements was looked upon as quite improbable before the advent of the Bohr theory.

The difference between the  $H$  and  $He^+$  wave-lengths is obtained by replacing  $M_{He}$  approximately by  $4M_H$ :

$$\frac{R_H}{R_{He}} = \frac{M_H(m + M_{He})}{M_{He}(m + M_H)} = \frac{1 + \frac{m}{4M_H}}{1 + \frac{m}{M_H}} \approx 1 - \frac{3}{4} \frac{m}{M_H}. \quad (34)$$

This shows that the wave-lengths of the hydrogen lines are 0.04 per cent greater than those of the neighbouring  $He^+$  lines. Reversing the procedure, we can use the very accurately measurable difference in wave-length to determine  $e/m$  spectroscopically, for we have

$$\frac{m}{M_H} = \frac{\frac{e}{M_H}}{\frac{e}{m}} = \frac{F}{\frac{e}{m}} = \frac{R_{He} - R_H}{R_H - \frac{1}{4}R_{He}}, \quad \text{i.e.} \quad \frac{e}{m} = F \frac{R_H - \frac{1}{4}R_{He}}{R_{He} - R_H}. \quad (35)$$

In the exact computation we must introduce—besides the precise atomic weight ratios—a series of corrections—e.g. the difference between atomic mass and nuclear mass must be taken into consideration.

Equation (31) is of importance in connexion with the discovery of deuterium, the hydrogen isotope of mass 2, whose nucleus is called a *deuteron*. Setting the doubled value in place of  $M_H$  in this equation yields the positions of the deuterium spectrum lines in excellent agreement with experiment.

## 5. X-ray Spectra. Moseley's Law.

An atom of atomic number  $Z$  which has lost all but one electron will emit radiations given by

$$\nu = Z^2 R \left( \frac{1}{n_f^2} - \frac{1}{n_i^2} \right). \quad \dots \quad (36)$$

if we neglect the motion of the nucleus, which, moreover, is of less importance as the atomic number increases. On account of the factor  $Z^2$ , atomic numbers as low as 30 bring these frequencies into the ordinary X-ray region. Highly ionized atoms of this kind cannot be produced

experimentally; nevertheless, there are X-ray lines which, in part, conform to the above law. In order to understand their origin we must examine in more detail our ideas concerning the structure of the heavier atoms. We have every reason to believe that even when such an atom is in its normal state all electrons are not in the lowest quantum state, corresponding to  $n = 1$ . Rather, there are groups of electrons having the same quantum number  $n$  which we may provisionally imagine to be arranged in various shells of radii  $a_1, a_2, \dots, a_n$ . These shells have come to be denoted by the letters  $K, L, M, N \dots$ , corresponding to the quantum numbers 1, 2, 3, 4,  $\dots$ . Now the X-ray spectra differ experimentally from the visible and ultra-violet spectra in that absorption *lines* do not occur. We observe merely continuous regions of absorption having abrupt terminal edges at the long wave-length end. This phenomenon can be explained as follows: in the case of visible spectra, the outermost electron is brought to a higher orbit in the process of absorption. If, however, one of the inner electrons—say a  $K$  electron—is to be raised to a higher level, we find that all the places in the  $L, M, \dots$  shells of a normal atom are occupied, so that this electron can find room only at the periphery of the atom. But the work necessary to bring it to that point is practically the same as that required for complete detachment, assuming the electron to possess no kinetic energy after removal. The long wave absorption limit, then, corresponds to the work done in detaching the inner electron, and the continuous spectrum which follows it corresponds to transitions in which the electron possesses varying amounts of kinetic energy after leaving the atom. Since the energy of translation of a free electron is not of a periodic nature it is not subject to a quantum condition of any kind, and so the energy levels form a continuous sequence. A similar continuous spectrum is found at the series limit of visible spectra, both in absorption and in emission. These continua originate in a similar way.

If, now, an inner electron is detached from a heavy atom by radiation or by electron impact, its place can be filled by an electron from another shell. This results in radiation. Neglecting all interactions between the electrons, we would find that the X-ray emission lines originating in this way follow the hydrogen formula. In any actual case it must be remembered that according to the laws of electrostatics the electrons in the inner shells screen the nuclear charge from the surroundings, so that we must introduce a screening constant  $s_n$  for each shell in such way that the energy of an  $n$ -quantum electron is given by

$$E_n = (Z - s_n)^2 \frac{Rhc}{n^2}. \quad \cdot \cdot \cdot \cdot \cdot (37)$$

The fact that  $s_n$  is in large measure independent of  $Z$  is a justification



of this somewhat crudely constructed theory. We can set  $s = 1$  for the  $K$  and  $L$  shells; then for the line corresponding to a transition from the  $L$  to the  $K$  shell we have

$$\bar{\nu} = (Z - 1)^2 R \left( \frac{1}{1^2} - \frac{1}{2^2} \right) = \frac{3}{4} (Z - 1)^2 R. \quad . \quad . \quad (38)$$

This is the celebrated Moseley Law for the  $K_\alpha$  line.\* This line is found in the X-ray spectra of all elements from  $Z = 4$  to  $Z = 92$ . The atomic number of any element can be determined at once with the help of Moseley's Law. However, the detailed theory of the X-ray terms is very complicated, and by no means yet complete. We have given here only the broadest outlines of the theory.

## 6. The Correspondence Principle.

Equation (17) (p. 648) states that the orbital frequency of the electron in a hydrogen atom is given by

$$\frac{\omega}{2\pi} = \nu_{cl} = \frac{4\pi^2 m e^4}{h^3 n^3}. \quad . \quad . \quad . \quad (17)$$

According to the classical connexion between radiation and electrical oscillations this is the same as the radiated frequency. On the other hand, the Quantum Theory gives the expression

$$\nu_{qu} = \frac{2\pi^2 e^4 m}{h^3} \cdot \frac{n_i^2 - n_f^2}{n_i^2 n_f^2} \quad . \quad . \quad . \quad (20)$$

for the frequency of the radiation connected with the transition  $n_i \rightarrow n_f$ . If  $n_i - n_f$  is small compared with  $n_i$  we can write instead

$$\nu_{qu} = \frac{4\pi^2 e^4 m}{h^3 n_i^3} (n_i - n_f). \quad . \quad . \quad . \quad (20')$$

We see that in the limiting case of large quantum numbers the transition  $\Delta n = 1$  yields a frequency identical with the classical frequency, the transition  $\Delta n = 2$  gives the first harmonic  $2\nu_{cl}$ , &c. Thus there is an asymptotic connexion between the frequencies given by the Quantum Theory and the classical frequencies corresponding to the mechanical motions of the system. For this reason a quantum theory computation of the radiation from an antenna would lead to the same result as that already obtained. The extension of this relationship, which is called the *Principle of Correspondence*, makes it possible to fill, in part, a gap in the previous theory. Thus, while the classical theory of the radiating oscillator permits a computation not only of

\* H. G. J. Moseley, *Phil. Mag.*, 26, p. 1024 (1913); *ibid.* 27, p. 703 (1914).

Since  $s_n$  shows a certain progression with  $Z$ , it is only the middle elements for which equation (38) is obeyed exactly.



the frequency but also of the intensity, polarization and coherence of the emitted radiation from a knowledge of what is occurring in the system (cf. p. 335 *et seq.*), the Bohr theory gives only the frequency. In the language of the Bohr theory, the intensity corresponds to the frequency of occurrence of a quantum transition. The two fundamental postulates have nothing to say concerning this quantity.

Bohr then assumes—and his assumption is justified by its success—that an asymptotic connexion obtains for intensity and polarization also, in such a way that the frequency of the transitions producing light of a given state of polarization is determined by the amplitude of the radiation as calculated on classical principles. For small numbers it is not determined whether the initial or the final orbit is to be taken—it is possible only to make an estimate. The results become unique, however, if a given harmonic component is absent from both the initial and the final states. In that case the corresponding quantum transition certainly does not occur. In this way we arrive at principles of selection which in many cases limit to a great extent the number of possibilities of combination of the various energy levels.

The Correspondence Principle also points the way to an extension of the quantum hypotheses to systems with several degrees of freedom. According to the classical theory, the frequencies of a system are given by (cf. p. 130)

$$\nu_1 = \frac{\partial E}{\partial J_1}, \quad \nu_2 = \frac{\partial E}{\partial J_2}, \quad \nu_3 = \frac{\partial E}{\partial J_3} \dots; \quad \cdot \cdot \quad (39)$$

hence any combination frequency is given by

$$\nu_{cl} = s_1 \frac{\partial E}{\partial J_1} + s_2 \frac{\partial E}{\partial J_2} + s_3 \frac{\partial E}{\partial J_3} \dots \quad \cdot \cdot \cdot \quad (40)$$

where the  $s_i$  are any positive or negative integers. Again, for a relatively small energy change, the energy difference is given as a function of the changes of the phase integrals. Neglecting higher terms in the Taylor expansion,

$$\Delta E = h\nu_{qu} = \frac{\partial E}{\partial J_1} \Delta J_1 + \frac{\partial E}{\partial J_2} \Delta J_2 + \frac{\partial E}{\partial J_3} \Delta J_3 \dots \quad (41)$$

If the asymptotic relation between the classical and the quantum frequency is to be retained, the phase integrals for the case of several degrees of freedom must change by integral multiples of  $h$ . This condition will certainly obtain if we impose the requirement that not only their changes but the quantities themselves are to be integral multiples of  $h$ . But it follows also from the asymptotic connexion that for degenerate systems there are only as many quantum conditions as the system has periods. The Kepler ellipse in space is a doubly-degenerate problem, for corresponding to the three co-ordinates of the electron

there is but a single frequency—that of the motion in the orbit. This means that the energy states of the Kepler ellipse depend upon only one of the action variables, viz.  $J_1$  (cf. p. 135). This determines only the major axis. Hence *all ellipses having the same major axis have the same energy*. The circular orbit also belongs to this set; therefore the formulæ of § 3 and § 4 are valid also for elliptic orbits, provided we understand  $a$  to mean the semi major axis. The introduction of elliptic orbits is necessary, since with only circular orbits the Correspondence Principle would yield the result that on account of the absence of overtones in the initial and final states only transitions for which  $\Delta n = 1$  would be allowed. This means that only the first member of each series would appear. For elliptic orbits, on the other hand, all higher harmonics are present.

### 7. The Atomic Spectra of the Alkalis, the Alkaline Earths and similar Systems.

All calculations of the energy levels of more complicated atoms start from the hydrogen terms, and the modifications of the term values are approximated to by calculating the perturbations due to the other electrons. In order to be able to perform this calculation, we require a far-reaching schematization of the atom, whose justification is, moreover, given by the newer Quantum Theory. We consider first a system which we have reason to believe, especially from chemistry, consists of a loosely-bound electron—the valence electron—and a core consisting of the remaining  $(Z - 1)$  electrons, and a nucleus with  $Z$  charges. The normal atoms of the first column of the periodic table—the alkalis and the series Cu, Ag, Au—are represented by such a system. Because of the screening effect of the core electrons, the valence electron is under the influence of an effective charge of one unit. If we neglect entirely the perturbations caused by the core, we obtain the hydrogen terms. Now it has been known for a long time that the alkali spectra have considerable similarity to the hydrogen spectrum. It is to be expected, then, that the deviations can be explained as disturbances caused by the electric field of the core. We seek to take this field into account by adding to the Coulomb potential of the effective unit nuclear charge a spherically symmetric perturbing potential in the form of a power series

$$U_s = \frac{c_2}{r^2} + \frac{c_3}{r^3} + \frac{c_4}{r^4} + \dots \quad (42)$$

Since the added field also represents a central force directed toward the nucleus, the orbit must remain plane, according to p. 89. Thus its orientation in space remains fixed and so the angle variable  $w_3$  of the Delaunay elements (p. 136) remains constant. However, the complete computation of this mechanical problem shows that the



perturbing potential causes an advance of the perihelion, i.e. the angle variable  $w_2$  becomes a true variable. As a result, the energy no longer depends on  $J_1 = J_r + J_\theta + J_\phi$  alone, but also on  $J_2 = J_\theta + J_\phi$ , and we have two independent fundamental frequencies: that of the motion in the ellipse and that of the advance of perihelion. On the basis of the Quantum Theory we are thus required to introduce, in addition to the principal quantum number  $n$  connected with the description of the ellipse, a secondary quantum number  $k$  in such way that  $J_2 = kh$ . Now it can be shown that  $J_2/J_1 = k/n$  gives the ratio of the minor to the major axis of the ellipse, so that we must have  $1 \leq k \leq n$ . The precessing ellipse traces a *rosette* orbit whose energy is found as follows. We compare this energy with that of the ellipse which osculates the true orbit at aphelion; this is the path which the electron would pursue if the perturbing field were absent. Since the potential of the core, and still more the field of the core, decrease with a comparatively high power of the distance, the potential energies of the two orbits may be taken to be approximately equal at aphelion. But the kinetic energies are also the same, for since the centripetal force is the same, the centrifugal force  $mv^2/\rho$  is also the same in both paths; moreover, the radius of curvature  $\rho$  is the same for the two osculating paths, so that the kinetic energies are identical. In general, the osculating ellipse will not be exactly a quantum orbit. We assign to it a *non-integral* effective quantum number  $n^*$  by means of the definition of hydrogen-like terms:

$$E = -\frac{Rhc}{n^{*2}} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (43)$$

Since, by the Law of Areas, the part of the orbit lying close to the core is traversed very rapidly, while the perturbations due to the core electrons are appreciable only along this part, we can—to a first approximation—set the rotational frequencies on the two paths equal to each other. For the ellipse, for which there is but one quantum number  $n^*$ , we have

$$\nu = \frac{dE(n^*)}{h dn^*}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (44)$$

while for the rosette we must write

$$\nu = \frac{\partial E(n, k)}{h \partial n}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (45)$$

since  $E$  depends also on  $k$ . From the equality of the two frequencies it follows that the difference  $(n - n^*)$  cannot depend upon  $n$  itself, but only upon  $k$ :†

$$n^* = n + \phi(k). \quad . \quad . \quad . \quad . \quad . \quad . \quad (46)$$

† For 
$$\frac{\partial E}{\partial n} = \frac{dE}{dn^*} = \frac{dE}{dn^*} \cdot \frac{\partial n^*}{\partial n}, \quad \text{i.e.} \quad \frac{\partial n^*}{\partial n} = 1.$$



We thus arrive at the "Rydberg Term"

$$E = \frac{-Rhc}{[n + \phi(k)]^2} \quad \cdot \cdot \cdot \cdot \cdot \quad (47)$$

We obtain various term sequences according to the value of  $k$ . That corresponding to  $k = 1$  is called the  $S$  sequence;  $k = 2$  yields the  $P$  sequence;  $k = 3$ , the  $D$  sequence;  $k = 4$ , the  $F$  sequence. The origin of this notation will be explained below.

A further improvement is obtained if, instead of setting the periods of ellipse and rosette equal, we assume that their difference depends only upon the secondary quantum number  $k$ . This assumption is justified by the fact that the perturbation takes place mainly on the inner portion of the orbit, and this is determined primarily by the eccentricity, i.e. by the secondary quantum number  $k$ . Thus is obtained the "Ritz Term"

$$E = \frac{-Rhc}{\left[ n + \phi(k) + \frac{\psi(k)}{n^{*2}} \right]^2} \quad \cdot \cdot \cdot \cdot \quad (48)$$

which represents each term sequence  $k = \text{const.}$  by means of two constants  $\phi(k)$  and  $\psi(k)$ . This formula yields an excellent representation of the observed data.

The possibilities of transition between the energy levels corresponding to various values of  $k$  are drastically limited by the Correspondence Principle. It is true that the Fourier analysis of the motion of a rotating Kepler ellipse yields all harmonics of the orbital frequency  $\nu_1$ , but on account of the uniform advance of perihelion no harmonics of the precessional frequency  $\nu_2$  result;  $\nu_2$  occurs only in the combination  $s\nu_1 \pm \nu_2$ . In terms of the Correspondence Principle this means that a change of the principal quantum number is always connected with a change of the secondary quantum number by  $\pm 1$ . This implies that an  $S$  term can combine only with a  $P$  term, a  $P$  term only with an  $S$  or a  $D$  term. In this way, we interpret the empirically found series for which the following notation is usual:

$$s = \phi(1), \quad p = \phi(2), \quad d = \phi(3), \quad f = \phi(4)$$

$$\sigma = \psi(1), \quad \pi = \psi(2), \quad \delta = \psi(3), \quad \phi = \psi(4)$$

$$nS = \frac{R}{(n + s)^2} \quad \text{or, more precisely,} \quad \frac{R}{\left[ n + s + \frac{\sigma}{n^{*2}} \right]^2}$$

$$nP = \frac{R}{(n + p)^2} \quad \text{or, more precisely,} \quad \frac{R}{\left[ n + p + \frac{\pi}{n^{*2}} \right]^2}$$

The series are then written simply:

Principal Series:  $\bar{\nu} = 1S - nP, \quad n = 2, 3, 4, \dots^*$

Sharp Series:  $\bar{\nu} = 2P - nS, \quad n = 2, 3, 4, \dots$

Diffuse Series:  $\bar{\nu} = 2P - nD, \quad n = 3, 4, 5, \dots$

Fundamental Series:  $\bar{\nu} = 3D - nF, \quad n = 4, 5, 6, \dots$

This terminology has the following origin. The Principal Series contains those lines which the normal atom can absorb, since the fixed term with which the  $P$  sequence combines is the ground term. These lines are also the easiest to excite in emission. The names of the Sharp and Diffuse series arise from insignificant differences in the sharpness of the lines. In the German literature these series are called the "Second Subordinate Series" and "First Subordinate Series" respectively. It is to be noticed that they have the same series limit. The Fundamental Series is so called because it was considered of particular significance on account of the fact that its terms differ but little from those of hydrogen. German writers refer to this series as the "Bergmann Series", after its discoverer. All possible combinations are not covered by these series, however. Apart from the lines fitting into the historic series, a large number of "combination lines", e.g.  $3D - 3P$ , are known.

Besides the spectra of neutral atoms of the first column of the periodic table, the spectra of singly-ionized alkaline earth metals and of doubly-ionized atoms of the third column are also of the alkali type. We note that in all these cases only one valence electron is present. The factor  $Z^2$  which must be supplied in the term formula is equal to 4 for the singly-ionized alkaline earths, and 9 for the doubly-ionized elements of the third column. But we can distinguish the same series also for the elements with two valence electrons, i.e. the neutral atoms of the second column.† We must imagine these series to originate when one of the two valence electrons occupies the orbits corresponding to the several energy levels, the other valence electron being counted as part of the core in computing the terms; i.e. for a spherically symmetric perturbing potential we again imagine its charge spread over a sphere. The fact that this approximation is better than might be expected is explained by the Wave Mechanics.

Besides the lines fitting into this system, it has been known for a long time that there are lines which originate by combination with a new sequence of energy levels empirically called the  $P'$  term sequence. The remarkable feature of these terms is that they become negative for higher values of the current number and converge to a negative

\* These current numbers are not the actual principal quantum numbers, but are the nearest integers to the effective quantum numbers.

† The difference between systems with one and with two valence electrons is manifested by the "fine structure" of the lines. This is discussed below.



limit. This evidently has the following meaning. If one of the valence electrons is raised to higher and higher  $P'$  levels and finally freed from the atom, the amount of work required exceeds the amount needed to remove it by way of the normal energy levels by the magnitude of the energy of the negative limit. But this work can be done only on the remainder of the atom, i.e. the limit of the  $P'$  terms corresponds to an ion with an excited electron. Thus the accented terms belong to atomic states in which both valence electrons are raised simultaneously. Then the difference  $\infty P - \infty P'$  must correspond to the difference between two terms of the ionized atom. As a matter of fact, for calcium, for example, this difference is equal to that of the  $1S$  and  $3D$  states of the ion.

The distribution of the energy of excitation among several electrons makes the spectra of elements with a number of equivalent electrons very complicated. The development of series becomes less and less pronounced, since the energy of excitation is employed to lift several electrons, instead of raising a single electron to a high orbit. However, the abundance of lines in such a spectrum is due in part also to a further difference which we shall examine in the following sections.

Before we take up these questions we shall attempt to give a more complete physical foundation for the potential of the core. If we compare the spectroscopically determined values of the functions  $\phi$  and  $\psi$  in the Ritz formula with those which would result mathematically from given coefficients  $c_2, c_3, c_4, \dots$ , we find that for the alkalis, from  $k = 3$  onward, the coefficient  $c_4$  of  $1/r^4$  predominates. This may be interpreted as follows. Under the influence of the valence electron, the electron shell will be displaced slightly in a direction away from the valence electron. As a result, the centre of gravity of the positive charge will no longer coincide with that of the negative charge, and a dipole is formed. If the polarizability of the ion is  $a$ , the dipole moment will be  $|m| = ae/r^2$ ; but its potential at the place occupied by the electron is

$$V = \frac{|m|}{r^2} = \frac{ae}{r^4} \quad \cdot \cdot \cdot \cdot \cdot \cdot (49)$$

Actually we obtain values of  $a$  from the Rydberg and Ritz corrections which agree well with refraction data on the ion corresponding to the core (cf. p. 452). For small values of  $k$ , i.e. for  $S$  and  $P$  terms, the Rydberg correction becomes so large that it can no longer be explained by polarization of the core. Schrödinger was able to show that large Rydberg corrections of this kind result when the valence electron penetrates into the region of the core electrons, which we imagine distributed over a spherical shell. Our derivation of the Rydberg formula makes no special assumptions as to the nature of the perturbing field, and so is valid for both cases.



## 8. Perturbation of Electron Orbits by External Forces. Stark Effect and Zeeman Effect.

The methods of Celestial Mechanics, whose fundamentals were derived in Chapter VI, Part II of this book (p. 108), together with the quantum conditions, permit us to calculate the perturbations of the term values caused by external electric or magnetic fields. The calculation becomes particularly simple when based upon two theorems which must be given here without proof.

The first theorem, the so-called Adiabatic Principle, states that if perturbations are produced in a system by slowly altering the external conditions, the energy is in general changed, but the phase integrals retain their former values. If, then, a state is once characterized as a quantum state by values of the phase integrals which are integral multiples of  $h$ , it remains so when the system undergoes an adiabatic change. The meaning of the theorem may be illustrated by means of the so-called Einstein pendulum. Let a pendulum bob suspended by a thread swing with an amplitude corresponding to a quantum state  $nh\nu$  of an oscillator. Let the frequency be raised to  $\nu'$  by slowly shortening the thread at the point of support. As may be verified readily in this case, the pendulum takes up enough energy to give it an amplitude corresponding to  $nh\nu'$ .

The second theorem, which becomes evident immediately, states that if there is a perturbing function in the expression for the energy, the value of the additional energy due to this cause is, to a first approximation, equal to the mean value of the perturbing energy taken over the unperturbed orbit. For example, if a hydrogen atom is brought into a homogeneous electric field  $F$  whose direction is that of the  $z$ -axis, a perturbation term  $-eFz$  is to be added to the Coulomb term  $-e^2/r$  in the expression for the energy, and the average change in energy is  $-eFz$ . Naturally, this depends on the eccentricity and orientation of the orbit, i.e. the orbits having the same major axis—which have the same energy in the unperturbed atom—assume different energy values when the field is applied; in other words, the energy levels are subdivided. This subdivision is very complicated in the case of hydrogen, and it must be looked upon as one of the finest accomplishments of the theory that it makes possible the computation of these patterns. This phenomenon is the *Stark Effect*, discovered independently by Stark and by Lo Surdo. The theory shows at once that the subdivision of the levels occurs in this form only for one-electron systems. For multiple-electron systems a splitting into orbits having various eccentricities but equal major axes is already caused by the field of the core. This is merely the decomposition into *S*, *P*, *D* term sequences. The external field can exert but a small additional influence which manifests itself in a displacement increasing

with the square of the field strength and, occasionally, by splitting. On the other hand, the field has a great influence on the *intensity* of the lines of these atoms; the field gives rise to harmonics which, according to the Correspondence Principle, make possible transitions which cannot occur in the absence of the field. Since very intense local fields occur in condensed spark discharges, we see why many "forbidden" lines appear under such conditions.

The perturbations caused by a magnetic field are of even greater importance in the investigation of the structure of the atom, for a field of this kind causes the spectral lines of the atoms to subdivide in very characteristic ways. This is the *Zeeman Effect*. Application of the second perturbation theorem leads into difficulty, since the force  $-e[\mathbf{v}\mathbf{H}]/c$  exerted by a magnetic field on an electron has no potential. One might even believe, at first, that a magnetic field could cause no change in energy at all, since the force is always normal to the path, and so can do no work on the electron. This is, of course, true for the field itself, but it must be remembered that while the magnetic field is building up there exists a time derivative  $\partial\mathbf{H}/\partial t$ , and hence, by the second Maxwell equation, an electric vorticity (cf. *Ex.* 89, p. 316). This curl determines an energy change, but at the same time the orbit remains a quantum orbit on account of the Adiabatic Principle. For simplicity, we shall assume the field to increase at a uniform rate, attaining its full value after  $N$  electronic rotations. Then we may replace  $\partial\mathbf{H}/\partial t$  in the Maxwell equation by  $\mathbf{H}/N\tau$ . The work done by the electric field during the time of one revolution of the electron is

$$\frac{W}{N} = -e \oint \mathbf{E} d\mathbf{s} = -e (\text{curl } \mathbf{E}) \mathbf{S} = \frac{e}{c} \frac{\mathbf{H}}{N\tau} \mathbf{S},$$

where  $\mathbf{S}$  denotes the area and orientation of the orbit. The total work done after  $N$  revolutions is then

$$W = \frac{e}{c} \frac{\mathbf{S}}{\tau} \mathbf{H}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (50)$$

On account of the Law of Areas we can replace  $\mathbf{S}/\tau$  by the areal velocity  $d\mathbf{S}/dt$  and express this in turn by the mechanical moment of momentum  $\mathbf{d}$ . Then we have

$$W = \frac{e}{2mc} \mathbf{d} \mathbf{H}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (51)$$

But according to p. 463, the magnetic moment of an electron orbit whose moment of momentum is  $\mathbf{d}$  is given by

$$\boldsymbol{\phi} = -\frac{e}{2mc} \mathbf{d}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (52)$$

so that

$$W = -\boldsymbol{\phi} \mathbf{H}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (53)$$



*The work done in increasing the field from zero to its final value  $\mathbf{H}$  is equal to the energy of the magnetic dipole of strength  $\mathbf{p}$  in the field  $\mathbf{H}$ .*

In what form is this energy present in the atom? Since there is no such quantity as potential energy of an electron in a magnetic field, this energy can be present only as kinetic energy. To a first approximation the change of motion can be given very simply. We prove a theorem, given by Larmor, to the effect that to a first approximation the only effect of the magnetic field is to cause a uniform rotation of the entire electron system about the field as an axis, the direction and magnitude of the rotation being given by

$$\boldsymbol{\omega} = \frac{e}{2mc} \mathbf{H}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (54)$$

If the co-ordinate system to which the electron motions are referred is in rotation with respect to an inertial system, the following inertial forces arise (cf. p. 233): the Coriolis force

$$\mathbf{F}_o = -2m[\boldsymbol{\omega} \mathbf{v}_{\text{rel}}], \quad . \quad . \quad . \quad . \quad . \quad (55)$$

and the centrifugal force

$$\mathbf{F}_r = -m[\boldsymbol{\omega}[\boldsymbol{\omega} \mathbf{r}]].$$

If the product  $[\boldsymbol{\omega} \mathbf{r}]$  is small compared with the velocity in the moving system—which is, of course, the case here—then the centrifugal force can be neglected in comparison with the Coriolis force. The force exerted by the magnetic field, on the other hand, is

$$\mathbf{F}_m = -\frac{e}{c} [\mathbf{v} \mathbf{H}], \quad . \quad . \quad . \quad . \quad . \quad (56)$$

where, strictly,  $\mathbf{v}$  should be replaced by the absolute velocity  $\mathbf{v}_{\text{rel}} + [\boldsymbol{\omega} \mathbf{r}]$ , but may be replaced approximately by  $\mathbf{v}_{\text{rel}}$ . The two forces neutralize each other, i.e. determine the same electron orbits in the moving system as those which would be present without the field in the system at rest, if

$$\mathbf{F}_o + \mathbf{F}_m = -\frac{e}{c} [\mathbf{v}_{\text{rel}} \mathbf{H}] + 2m[\mathbf{v}_{\text{rel}} \boldsymbol{\omega}] = 0, \quad \text{or} \quad \boldsymbol{\omega} = \frac{e}{2mc} \mathbf{H}. \quad (57)$$

This is known as the Larmor precession. The corresponding change in kinetic energy—which may be positive or negative according to the relative positions of  $\boldsymbol{\omega}$  and the orbital angular momentum—is the work done by the field. In what follows we shall use the convenient formula (53) to determine the change of the term values.

As long as any arbitrary orientation of  $\mathbf{p}$ —or, what amounts to the same thing, of the angular momentum vector  $\mathbf{d}$ —with respect to the field  $\mathbf{H}$  is permissible, we obtain a continuous manifold of energy



changes. This means that a quantized energy value would become diffuse in the field, so that no splitting of lines but only a broadening could take place in a magnetic field. However, it must be remembered that the Larmor precession of a Kepler ellipse determines a uniform rotation of the line of nodes, and the third Delaunay variable  $w_3$  becomes a true variable (cf. p. 136). Then, according to § 6 (p. 656), the associated action variable  $J_3$  must become an integral multiple of  $h$ . This quantity represents  $2\pi$  times the  $z$ -component of the moment of momentum. Since  $2\pi$  times the total angular momentum is itself an action variable  $J_2$ , which is an integral multiple of  $h$ , it follows that if we set  $J_3 = mh$  the inclination of the orbit is given by

$$\cos i = \frac{J_3}{J_2} = \frac{m}{k}. \quad \cdot \cdot \cdot \cdot \cdot \cdot (58)$$

We thus arrive at a directional quantization. For  $k = 1$  only the values  $m = +1, 0, -1$ , i.e. the angles  $0, \pi/2, \pi$  are possible. It is to be noticed that the quantum numbers  $n$  and  $k$  are essentially positive quantities, but the "magnetic" quantum number  $m$ , on the other hand, can assume both positive and negative values since an external field determines a definite sense as well as direction.

This directional (or spatial) quantization, which is of especial importance in the Theory of Paramagnetism, was directly verified by the experiments of Stern and Gerlach.\* In these experiments a fine pencil of atoms of silver or some other metal was produced by evaporating the metal in a high vacuum. The beam was sent through a non-homogeneous magnetic field whose gradient  $\text{grad } |\mathbf{H}|$  had practically the same direction as  $\mathbf{H}$ . Such a field exists, for example, between two pole pieces, one of which is a flat plate, the other a sharp edge.† A magnetic dipole experiences a translational force in such a field (cf. *Ex.* 78, p. 286); this force is in the direction of  $\mathbf{H}$  and is given by

$$|\mathbf{F}| = |\boldsymbol{\phi}| |\text{grad } \mathbf{H}| \cos(\boldsymbol{\phi} \mathbf{H}). \quad \cdot \cdot \cdot (59)$$

If, as would be expected in the classical statistics, all orientations of  $\boldsymbol{\phi}$  occurred as a result of the thermal motion, the beam would experience a continuous broadening in the magnetic field. In reality, however, the beam of silver atoms was found to break up into two parts, corresponding to the orientations  $0$  and  $\pi$ . There is a particular reason for the absence of the orientation  $\pi/2$  (cf. p. 670).

The Quantum Theory provides a fundamental unit for the magnitude of the magnetic moment, the *Bohr Magneton*. Since the total angular momentum is an action variable and so must be an integral

\* O. Stern and W. Gerlach, *Ann. der Physik*, 74, p. 673 (1924).

† This type of field is not strictly of zero divergence. In carrying out the experiment it is even more closely approximated by means of a groove of rectangular cross-section that is cut in the plate.

multiple of  $h/2\pi$  (this holds for all atoms, not only for hydrogen), it follows from the magneto-mechanical parallelism (p. 463) that the magnetic moment must also be an integral multiple of the quantity

$$\mu_B = \frac{eh}{4\pi m_e c} = 9.27 \times 10^{-21} \text{ e.m.u.} \quad . \quad . \quad . \quad (60)$$

The atomic ray experiments actually gave a deviation for silver which yielded  $|\boldsymbol{\phi}| = 1 \cdot \mu_B$ , but one cannot expect the susceptibility measurements evaluated according to p. 458 to give integral numbers of magnetons, since the directional quantization is not assumed in this evaluation (cf. p. 672).

With the aid of equation (53) (p. 663) and equation (60), we can write at once the energy changes experienced in a magnetic field by an atom with secondary quantum number  $k$ , i.e. with moment of momentum  $kh/2\pi$ :

$$\begin{aligned} \Delta E = -\boldsymbol{\phi} \mathbf{H} &= +\mu_B k H \cos i = \pm \mu_B H m \\ &= \pm \frac{ehH}{4\pi m_e c} m, \quad 0 \leq |m| \leq k. \quad . \quad (61) \end{aligned}$$

In order to find how the spectral *lines* are split up, we must derive selection principles for the spatial quantum number  $m$ . We again turn to the Correspondence Principle and use it to determine the nature of the light observed in various directions when the radiation takes place according to the classical theory. We can determine the time rates of change of the components of the dipole moment from the motions of the electrons. By p. 335 *et seq.* these quantities determine the radiation of the "antenna system". At the same time, we must bear in mind the fact that a Hertzian oscillator emits no radiation in the direction of its axis. Hence, if we observe in the direction of the magnetic field ("longitudinal effect") we do not perceive the  $z$ -component of the motion of the electrons, which is not affected by the field. The  $x$ - and  $y$ -components, which execute the Larmor precession, are observed completely. The Fourier analysis yields combination frequencies of the form  $s\nu_1 \pm \nu_2 \pm \nu_3$ . In accordance with the Correspondence Principle, this means that only the quantum transition  $\Delta m = \pm 1$  occurs for radiation in the direction of the field, and that since the classical radiation is circularly polarized in the initial and final orbits, the emitted light is circularly polarized. In observing perpendicular to the field ("transverse effect"), the uninfluenced component is observed completely. Its Fourier analysis does not contain  $\nu_3$ , but only frequencies of the form  $s\nu_1 \pm \nu_2$ , so that we have quantum transitions  $\Delta m = 0$  connected with radiation normal to the field. The light vibrations are then parallel to the field. Of the transverse components of motion, i.e. the  $x$ - and  $y$ -components, the one in the line of sight will not be observed; the light sent out by



it is linearly polarized in a direction normal to the field. Since the combinations  $\nu_1 \pm \nu_2 \pm \nu_3$  again occur in the Fourier analysis, the transitions  $\Delta m = \pm 1$  correspond to them. Then, on account of the fact that  $m_i - m_f$  is limited to  $\pm 1$  or 0, the lines are divided in the following way:

$$\Delta\nu = \frac{ehH}{h4\pi m_e c} (m_i - m_f) = 0 \quad \text{or} \quad \Delta\nu = \pm \frac{eH}{4\pi m_e c}. \quad (62)$$

In the longitudinal effect the value 0 is excluded and the light of both components is either right-handed or left-handed circularly polarized,\* while in the transverse effect the undisplaced line is polarized in a direction parallel to the field, the side components in a direction normal to the field. This pattern of lines is called the *normal Lorentz Triplet*. The theory of this pattern was derived by Lorentz using the simple linear oscillator as a model of the radiating atom.† The Quantum Theory goes no farther; this is strange, since the normal Zeeman effect is shown by relatively few lines, and most lines exhibit much more complicated patterns containing large numbers of components—the so-called *anomalous Zeeman effect*. Nevertheless, regardless of how complicated the atom may be, the total angular momentum being constant in time is an action variable, and hence subject to the quantum condition; thus our formula is generally valid as long as the normal connexion between moment of momentum and magnetic moment obtains.

#### 9. Difficulties arising in the Atomic Theory of Magnetism. Explanation in Terms of the Spinning Electron.

The anomalous Zeeman effect is not the only difficulty offered by magnetism in the mechanics of the atom. The fine structure of many lines, e.g. the division of the familiar yellow line of sodium  $1S - 2P$  into two components  $\lambda 5890$  and  $\lambda 5896$  A.U. even in the absence of a field, is still unexplained. The components of this multiplet are really connected, and are not formed merely by lines which happen to lie near each other: if these lines are examined in a magnetic field of increasing strength, the pattern changes when the separation of the Zeeman components becomes comparable with that of the original lines and passes over into a single normal Lorentz triplet for very strong fields. This is known as the *Paschen-Back Effect*. One would conclude, then, that the doublet is caused by an internal magnetic field, and the location of this field would be assumed to be the core. But this cannot be, for the alkali ions corresponding to the core are known to be diamagnetic, and so are incapable of generating a field.

\* It is convenient to remember that the *short-wave* component  $\nu + \Delta\nu$  rotates in the same sense as the current which would cause the external field.

† H. A. Lorentz, *The Theory of Electrons*, London, Macmillan & Co. (1909).



Atoms with several valence electrons exhibit multiplets of much greater complexity. According to our previous computations these components should all come together to form a single line. For simplicity we consider not the splitting of the *lines*, which are always given by the difference of two terms, but rather that of the terms themselves. The *S* term is always single. The multiplicity increases with *k* up to a maximum value and remains constant thereafter; the multiplicity does not depend on the principal quantum number, and so is a characteristic quantity for each term sequence. The term system is designated as singlet, doublet, triplet, . . . according to the highest multiplicity. This multiplicity is written as a superscript to the left of the term symbol; thus  $^3S$  signifies the "triplet *S* state" of the atom. An atom with several valence electrons has several term systems with various multiplicities. For example, the alkaline earths possess a singlet system and a triplet system. As we proceed from column to column of the periodic table the even and odd multiplicities alternate. The multiplicities of a singly-ionized atom correspond to those of the preceding neutral atom. The highest multiplicity in the term system of an atom is greater by one than the number of valence electrons, i.e. one larger than the number of the column of the periodic table in which the atom belongs. Thus, for an element of column VII—e.g. manganese—octuple terms can occur. Were it not for selection rules, 64 lines could result from combinations between two such octet levels. Despite the fact that this number is reduced by the existence of selection rules, the tremendous richness of the spectra of atoms with many equivalently bound electrons is easy to understand. An analysis, however, yields a representation of the spectrum as combinations of a moderate number of multiplet terms. The lines resulting from the division of the terms may be represented formally by giving each sub-level a new quantum number *j*, introduced by Sommerfeld as the "*inner*" quantum number. The selection rule for *j* is

$$\Delta j = 0 \quad \text{or} \quad \pm 1, \text{ with the transition } 0 \rightarrow 0 \text{ forbidden.}$$

Concerning the physical significance of *j*, Sommerfeld surmised at the very first that  $j\hbar/2\pi$  represents the total angular momentum of the atom; this may now be considered a proven fact. For an alkali atom, whose core is not magnetic and which therefore has no angular momentum, *j* would then appear to be identical with *k*, the angular momentum of the orbital motion of the valence electron. But this is evidently not the case, for there are two different values of *j* corresponding to the same *k* in the *P* term; hence there must be something else in the atom which is magnetic and which possesses angular momentum.

A further remarkable fact is the following: The spectra of the one-

electron systems  $H$  and  $He^+$  also exhibit "fine structure" which, however, is quantitatively explained by a theory taking into account the relativistic variability of mass of the electron. Even in the absence of perturbations by the core, the orbits of various eccentricities have different energies on this account. For the alkalis, the splitting due to the different eccentricities is greatly increased by the core potential and the orbits of different eccentricity are separated into  $S$ ,  $P$ ,  $D$  terms, so that the relativistic effect plays a comparatively insignificant part. Nevertheless, the alkali doublets follow the same law as the relativistic hydrogen doublets as far as the dependence on nuclear charge is concerned. This might lead one to abandon the magnetic interpretation of the alkali doublets, yet there is still the possibility that both types of division follow the same law as a matter of chance.

Finally, we refer in this connexion to the magneto-mechanical anomaly found in the Einstein-de Haas Effect (p. 463), in which values of the magnetism of ferromagnetic substances result which are twice as great as those to be expected according to the magneto-mechanical parallelism.

All of these difficulties were removed at one stroke by Goudsmit and Uhlenbeck by their bold assumption that *the electron itself possesses a magnetic moment and a mechanical angular momentum*.\* The magnetic moment is assumed to be one Bohr magneton, while the moment of momentum is taken to be only  $1/2(h/2\pi)$ . These assumptions are justified by their results. The magnetic moment may be pictured as originating from rotation of the electronic charge about an axis, and it is not difficult to find the distribution of charge which gives the correct ratio of magnetic moment to angular momentum. However, no deeper meaning is to be attached to such a detailed picture of the structure of the electron, since it cannot be verified in any way. Nevertheless, this concept led to the term *spinning electron*. We refer to the inherent angular momentum as the *spin* of the electron. It is now understood that the magneto-mechanical anomaly for ferromagnetic substances—which are, without exception, metals—is due to the conduction electrons, while for salts—where the magnetism is the resultant of the orbital and spin moments—the factor lies between 1 and 2.

#### 10. The Theory of Multiplets and of their Zeeman Effects. Quantum Theory of Paramagnetic Susceptibility.

With the aid of the concept of electron spin one can readily bring order out of the chaos of multiplets. We limit our considerations to so-called normal multiplets for which the interaction of the electrons—called  $LS$  or Russell-Saunders coupling—may be represented by the following vector construction:

1. All spin vectors  $s_i$  may be combined *algebraically* into a resultant

\* S. Goudsmit and G. Uhlenbeck, *Zeitsch f. Phys.*, 35, 8-9, p. 618 (1926).



spin  $s$  which is an integer or a half-integer (measured in units of  $h/2\pi$ ):

$$s = \sum s_i = \sum \pm \frac{1}{2}.$$

Thus for two electrons the resultant spin is either 0 or 1.

2. All orbital angular momenta are to be combined *vectorially* to yield an integral resultant. Here we must introduce a modification which shows, for the first time, the insufficiency of our mechanical models: In order to keep the results in harmony with experience we find it necessary to decrease the orbital angular momenta by one unit. If these were the actual momenta, it would mean that the  $S$  orbits possess zero angular momentum, and would thus be straight lines through the nucleus. Only the newer quantum mechanics, which abstains from all special models, is capable of leading compulsorily to this numeration. One less than the secondary quantum number for

each electron is called  $l_i$ . Thus  $l_i = k_i - 1$ . The integral resultant for several electrons is called the group quantum number  $l$ . The symbols  $S$ ,  $P$ ,  $D$  will be applied to the group quantum numbers, while the corresponding small letters will be used for the individual electrons.

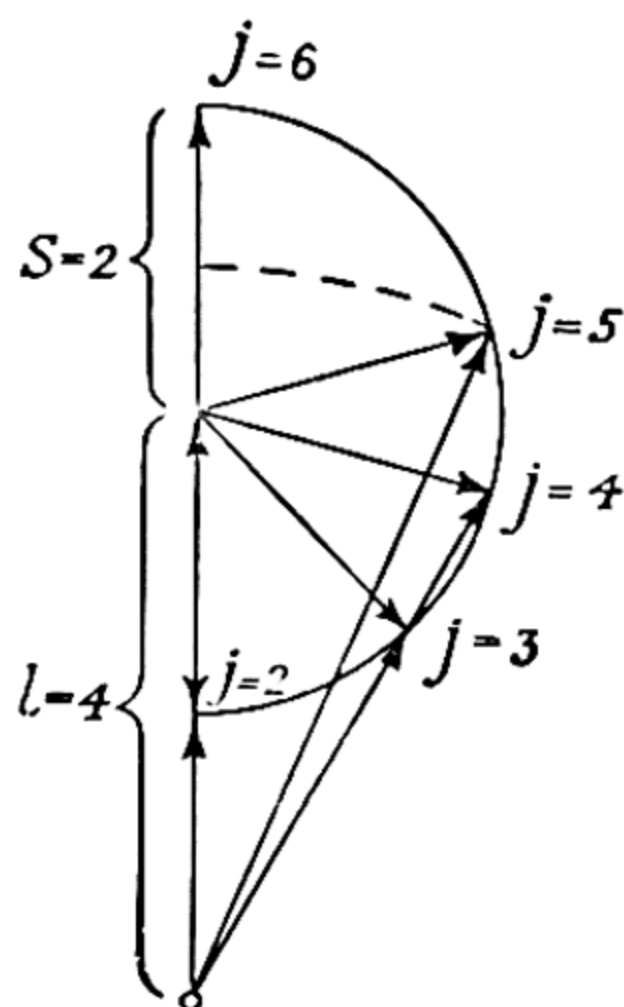


Fig. 3

3. The resultants  $s$  and  $l$  are to be combined *vectorially* into a resultant  $j$  which is integral (including zero) or half-integer, according as  $s$  is one or the other. The various integral or half-integer resultants are found by means of the following construction. Assume  $l > s$ . Describe a circle of radius  $s$  about the terminus of the vector  $l$  (fig. 3). The terminus of the resultant must lie on this circle. Its in-

tegral or half-integer values are constructed by drawing circles about the initial point of  $l$ , using the integers and half-integers as radii. The intersections of these circles with that of radius  $s$  mark the possible end points of the resultant vector  $j$ . It is seen that for  $l > s$  the number of possibilities, for integral or half-integer spin resultant, is  $2s + 1$ . This is then the multiplicity  $\tau$  of a term system corresponding to the spin resultant  $s$ . Since the latter is at most  $N/2$  when there are  $N$  valence electrons, the maximum multiplicity is  $N + 1$ . The other empirical rules given on p. 668 are obtained with equal ease from this construction. We attach the quantum number  $j$  to the term symbol as a subscript, e.g.  $2^2P_{1/2}$ .

4. In a weak external magnetic field (additional magnetic energy small compared with the natural multiplet separation) the vector  $j$  sets itself with respect to the field in such a way that its projection on the field direction is a positive or negative integer or half-integer



according as  $j$  itself is one or the other. This gives  $(2j + 1)$  positions, since the projection 0 is allowable only when  $j$  is integral.

5. In a strong magnetic field, on the other hand, the coupling between the orbital and spin vectors is dissolved; the resultant orbital angular momentum orients itself in such manner that its projection in the direction of the field is a whole number and, similarly, the resultant spin sets itself so that its projection on the field is a half-integer.

These five rules not only embody a correct representation of multiplets, but also lead to a formula for their Zeeman effect which is almost exactly correct.

If we now compute the additional energy due to the magnetic field by means of the expression  $-\boldsymbol{\mu} \cdot \mathbf{H}$ , we are not permitted to assume the normal connexion between the total angular momentum  $j\hbar/2\pi$  and the magnetic moment, for  $j$  is a combination of the normal orbital angular momentum with the anomalous spin. The true relationship is obtained from the following considerations: the invariable axis of the atom is given by the direction of the total angular momentum vector  $j$ . The vector of the resultant orbital angular momentum  $L$  precesses about this line. Since  $j$  is by construction the resultant of  $L$  and  $s$ , this means that the entire triangle  $OLS$  turns about its longest side. On the average, then, only the components of the vectors  $L$  and  $s$  in the direction of  $j$  contribute to the magnetic moment. On account of the anomaly of the spin, the component of  $s$  must be included twice, so that the magnetic moment becomes (cf. fig. 4)

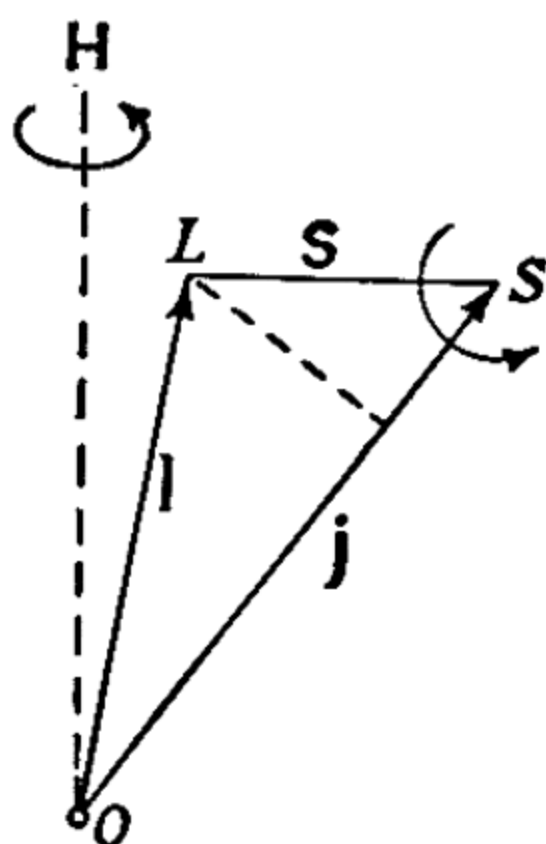


Fig. 4

$$|\boldsymbol{\mu}| = \frac{eh}{4\pi m_e c} [j + s \cos(sj)], \quad \dots \quad (63)$$

and since

$$\cos(sj) = \frac{s^2 + j^2 - l^2}{2sj}, \quad \dots \quad (64)$$

we have

$$|\boldsymbol{\mu}| = \frac{ehj}{4\pi m_e c} \left( 1 + \frac{s^2 + j^2 - l^2}{2j^2} \right) = \mu_B jg. \quad \dots \quad (65)$$

The factor in brackets,  $g$ , which measures the ratio of the actual magnetic moment to the one to be expected in the normal case, is called the *gyromagnetic ratio*. The above value, however, does not yet correspond with experiment. The strange thing is that the correct formula is obtained only after replacing each  $x^2$  by  $x(x + 1)$ , i.e.

$$g = 1 + \frac{j(j + 1) + s(s + 1) - l(l + 1)}{2j(j + 1)}. \quad \dots \quad (66)$$

This again demonstrates that the relations derived from models reproduce the facts approximately, but not absolutely correctly. It is only the new Quantum Theory which is capable of justifying formula (66).

The fact that  $g$  has different values in the initial and final states gives rise to the large number of components in the multiplet pattern. It follows from the energy increments

$$\Delta E_i = \mu_B g_i m_i H \quad \text{and} \quad \Delta E_f = \mu_B g_f m_f H \quad . \quad . \quad (67)$$

that the  $\pi$  components of the transverse effect (polarized with the electric vector parallel to  $\mathbf{H}$ ), for which  $m_i = m_f = \pm m$ , are given by

$$h\Delta\nu_\pi = \pm \mu_B H m (g_i - g_f), \quad . \quad . \quad . \quad (68)$$

and the  $\sigma$  components of the transverse effect (polarized with the electric vector perpendicular to  $\mathbf{H}$ ), for which  $m_i = m \pm 1$ ,  $m_f = \pm m$ , are given by

$$h\Delta\nu_\sigma = \pm \mu_B H [m(g_i - g_f) \pm g_i]. \quad . \quad . \quad (69)$$

As we see, the normal pattern is obtained for  $g_i = g_f = 1$ .

If, however, the magnetic field is *very strong*, the resultant orbital and spin angular momenta orient themselves independently, and we have the energy increment given by

$$\Delta E = \mu_B m_l H + 2\mu_B m_s H, \quad . \quad . \quad . \quad (70)$$

where  $m_l$  and  $m_s$  are the projections of  $\mathbf{l}$  and  $\mathbf{s}$  on the direction of  $\mathbf{H}$ ; and since  $m_s$  can have only the values  $+1/2$  or  $-1/2$ , the energy increment is that found in the normal effect.

On the basis of these considerations the calculation of magnetic susceptibility performed on p. 458 must be revised also. There we found that the magnetization of unit volume was given, according to the classical statistics, by

$$\mathbf{M} = \frac{Np^2\mathbf{H}}{kT} \overline{\cos^2 \theta}.$$

Referring all quantities to one mol, the susceptibility then becomes

$$\chi = \frac{(Lp)^2}{RT} \overline{\cos^2 \theta}. \quad . \quad . \quad . \quad (71)$$

Since  $\overline{\cos^2 \theta} = 1/3$  in the classical statistics, we have a unique connexion between the susceptibility and the atomic moment  $p$ . According to quantum principles,

$$p = gj\mu_B, \quad . \quad . \quad . \quad (72)$$

and since only  $2j + 1$  orientations—viz. those for which the magnetic

quantum number  $m$  is a whole number—are allowable, the mean value  $\overline{\cos^2 \theta}$  becomes

$$\overline{\cos^2 \theta} = \frac{1}{2j+1} \sum_{+j}^j \frac{m^2}{j^2} = \frac{2j(j+1)(2j+1)}{(2j+1)j^2 \cdot 6} = \frac{1}{3} \frac{j+1}{j}. \quad (73)$$

This formula is valid for integral as well as half-integral values of  $j$ , as one may verify readily. Then we have

$$\chi = \frac{L^2 g^2 j(j+1) \mu_B^2}{3RT}. \quad \cdot \cdot \cdot \cdot \cdot \quad (74)$$

On account of (72), the Curie constant, which was  $L^2 p^2 / 3R$  in the classical case, is now

$$C = \frac{L^2 g^2 j(j+1) \mu_B^2}{3R} = \frac{L^2 p^2 (j+1)}{3R j}. \quad \cdot \cdot \cdot \quad (75)$$

In order to calculate the elementary magnetic moment  $p$  from the observed value of  $C$  it would be necessary to know the value of  $j$  corresponding to the atomic state in question. One must therefore proceed in the opposite way by calculating  $g$  and then  $C$  from the values of the quantum numbers  $l$ ,  $s$  and  $j$  determined in some other way—spectroscopically or purely theoretically—and then compare the results with observation. A test of (75) which is free from objection must be undertaken by means of susceptibility measurements on paramagnetic, monatomic gases, since it is only here that disturbing influences are excluded. Unfortunately, the experimental obstacles have not yet been overcome, so that one must turn to experiments on solutions containing paramagnetic ions. Among the ions of the rare earths we find very good agreement between theory and experiment; this is understandable because, unlike the ions of the iron group, the electron shell responsible for paramagnetism is not at the surface and so is not subject to disturbances from neighbouring molecules. These disturbances prevent application of (75) in the case of the iron group, but at the same time they make conditions simpler. To a first approximation take all the terms of the ions of the iron group to be  $S$  terms. In this instance equation (75) becomes

$$C = \frac{4L^2 S(S+1)}{3R} \mu_B^2; \quad \cdot \cdot \cdot \cdot \cdot \quad (75')$$

i.e. the effective number of Bohr magnetons becomes

$$p_{eff} = \sqrt{4S(S+1)}. \quad \cdot \cdot \cdot \cdot \cdot \quad (75'')$$

The  $LS$  coupling treated here is not always developed to the exclusion of everything else. Among the heavier elements, transition types, representing the analogue of an incipient Paschen-Back effect, appear. Here we encounter for the



first time the general principle that the orientation of the spin vectors conforms to the magnitude of the energy of interaction. In a weak field the coupling between  $l$  and  $s$  overshadows that between the magnets and the field, so that  $l$  and  $s$  are first combined to produce  $j$ , which then orients itself in the field. But in a strong field both the coupling between  $l$  and the field, and that between  $s$  and the field, exceed that between  $l$  and  $s$ . By the same principle the  $LS$  coupling is effective only when the energy of coupling between the orbit and spin of an electron is small compared with the magnitudes of these energies among the several electrons. The spin-orbit coupling energy increases rapidly with the effective nuclear charge, so that the conditions no longer apply to heavy atoms, particularly those with many valence electrons. The other extreme case, called  $jj$  coupling, is encountered, for example, in the spectra of multiply ionized lead. In this case the spin and orbital momentum of an individual electron combine into a vector  $j_i = l_i \pm \frac{1}{2}$ . The total angular momentum is the integral or half-integral vector sum of the  $j_i$ , but there is no resultant  $l$  or  $s$ , so that the nomenclature based on  $LS$  coupling cannot be applied at all. If this is occasionally done in connexion with, say,  $Pb$  terms, it is by analogy with the terms of homologous systems having smaller nuclear charge. X-ray terms, too, are of the  $jj$  type. Of course the  $g$  formula becomes quite different for  $jj$  coupling. A different  $g$  value must be substituted in the first part of (75), but the second part remains unchanged because it contains only the magnetic moment and the quantum number  $J$ , which is valid in both cases.

*Ex. 132.* Sketch the relative positions of the terms arising in the case of two electrons with  $l_1 = 1$  and  $l_2 = 0$  for (a)  $LS$  coupling, (b)  $jj$  coupling.

*Ex. 133.* The *Faraday Effect* is the rotation of the plane of polarization of a beam of light when passed through a dense transparent medium situated in a magnetic field, the ray being directed along the field. Develop an approximate theory of this effect by combining the results obtained for the Zeeman Effect with the dispersion formula of p. 453.

## 11. The Structure of the Periodic System of the Elements. Pauli's Principle.

After arriving at the concept that the electrons of the more complicated atoms are arranged in shells having various quantum numbers, it was natural to assume that a new shell is begun with each period, i.e. each time we come to an alkali metal. The principal quantum number of the valence electron would therefore have the values Li 2, Na 3, K 4, Rb 5, Cs 6. Hence the lowest term of Na, which corresponds to the normal state, must be given the principal quantum number 3. But there is evidence against this, for the effective quantum number  $n^*$  obtained from the spectroscopic term value lies in the neighbourhood of 1.5. There must therefore be a very large Rydberg correction. That this is really the case is shown by the following considerations: Let us examine the similar systems  $Mg^+$ ,  $Al^{++}$ ,  $Si^{+3}$ ,  $P^{+4}$ ,  $S^{+5}$  \* in which there is likewise one valence electron and a core consisting of ten electrons. It is to be expected that as the effective nuclear charge—which amounts to 2 for  $Mg^+$ , 3 for  $Al^{++}$ , &c.—increases, the per-

\* Fowler introduced another notation for the spectra of atoms in various stages of ionization: the "arc", "spark" and "super-spark" spectra of an element  $E$  are referred to as  $E_1$ ,  $E_2$ ,  $E_3$ , &c., in this order.

turbations caused by the core will decrease. If we plot the values of the  $1S$ ,  $2P$  and  $3D$  terms as functions of the atomic number, at the same time dividing each term by  $Z^2$  in order to bring them to the same scale, we see from fig. 5 that both the  $S$  and the  $P$  terms converge toward the value of the three-quantum hydrogen level. Thus for the sodium atom the perturbation (caused by penetration of the orbit into the core) is really as great as indicated, and the above assignment of principal quantum numbers for the valence electron of the alkalis is justified.

We follow Bohr in assuming an atom to be built up by adding electrons successively to a bare nucleus. It is then natural to assume that the quantum numbers of the orbits of the first, second, third, &c., electrons are independent of  $Z$ , so that a system containing  $N$  electrons would be a model of the atom of atomic number  $N$ .\* This rule, suggestive of the Fundamental Law of Biogenetics, is in general obeyed, but is violated at certain places in the periodic table, i.e. for certain values of  $Z$ , for which a reconstruction takes place. That is, if we consider the spectroscopically determined quantum numbers corresponding to the lowest terms—the normal states—of the atom, we find that the electrons in the first row are in  $s$  orbits. There is indeed no other possibility when  $n = 1$ , since  $k \leq n$  or, otherwise expressed,  $l \leq n - 1$ . In the second row we find, besides  $s$  orbits,  $p$  orbits corresponding to the principal quantum number 2. The ground states of the atoms of the third row—the elements of which correspond both chemically and spectroscopically to those immediately above them—also have only  $s$  and  $p$  orbits, although, *a priori*,  $d$  orbits are also to be expected for  $n = 3$ . In Row 4 there are certainly  $4s$  orbits for K and Ca. Now the following elements, Sc to Cu, exhibit spectroscopic and chemical behaviour quite different from that of the elements of the preceding row. Spectroscopically they are distinguished by high multiplicities of the terms, while in the previous rows the high multiplicities possible according to the number of valence electrons (i.e. the column number) are not observed. The elements Sc to Cu are charac-

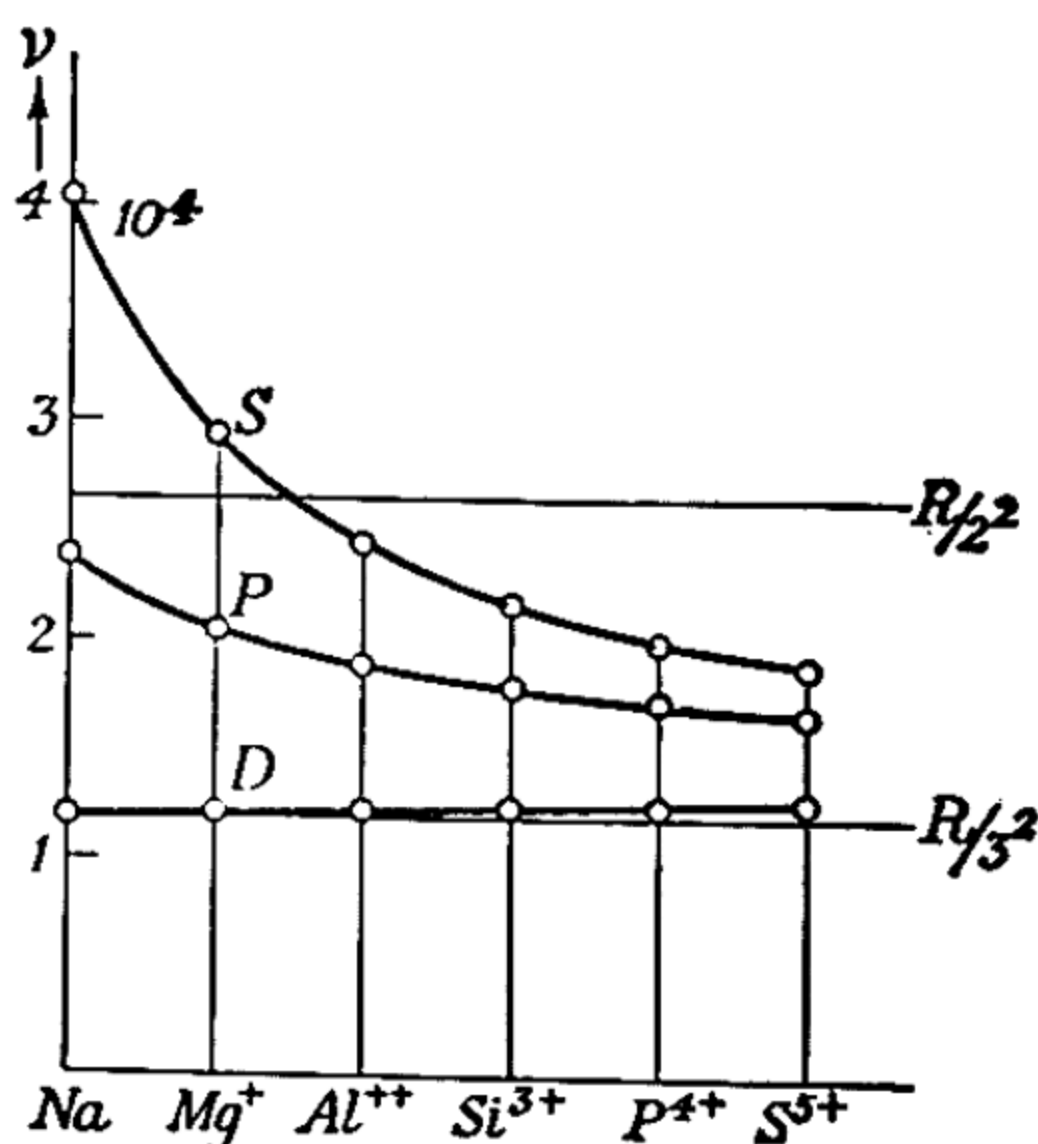


Fig. 5

\* This model would, of course, be of a reduced size because of the greater nuclear charge.



# PERIODIC TABLE OF THE ELEMENTS

<i>n</i>	I		II		III		IV		V		VI		VII		VIII		
	<i>a</i>	<i>b</i>	<i>a</i>	<i>b</i>	<i>a</i>	<i>b</i>	<i>a</i>	<i>b</i>	<i>a</i>	<i>b</i>	<i>a</i>	<i>b</i>	<i>a</i>	<i>b</i>	<i>a</i>	<i>b</i>	
1	1H 1.008															2 He 4.003	
2	3 Li 6.94		4 Be 9.02		5 B 10.82		6 C 12.01		7 N 14.01		8 O 16.00		9 F 19.00			10 Ne 20.18	
3	11 Na 23.00		12 Mg 24.32		13 Al 26.97		14 Si 28.06		15 P 31.02		16 S 32.06		17 Cl 35.46			18 A 39.94	
3 <i>d</i>	19 K 39.10		20 Ca 40.08		21 Sc 45.10		22 Ti 47.90		23 V 50.95		24 Cr 52.01		25 Mn 54.93		26 Fe 27 Co 28 Ni 55.85 58.94 58.69		
4	29 Cu 63.57		30 Zn 65.38		31 Ga 69.72		32 Ge 72.60		33 As 74.91		34 Se 78.96		35 Br 79.92		36 Kr 83.7		
4 <i>d</i>	37 Rb 85.48		38 Sr 87.63		39 Y 88.92		40 Zr 91.22		41 Nb 92.9		42 Mo 95.95		43 Tc 99		44 Ru 45 Rh 46 Pd 101.7 102.9 106.7		
5	47 Ag 107.88		48 Cd 112.41		49 In 114.76		50 Sn 118.70		51 Sb 121.76		52 Te 127.61		53 I 126.92		54 X 131.3		
5 <i>d</i>	55 Cs 132.91		56 Ba 137.36		57 La <span>[4f]</span> 138.92		72 Hf 178.6		73 Ta 180.88		74 W 183.92		75 Re 186.31		76 Os 77 Ir 78 Pt 190.2 193.1 195.23		
6	79 Au 197.2		80 Hg 200.61		81 Tl 204.39		82 Pb 207.21		83 Bi 209.0		84 Po 210.0		85 At 211		86 Rn 222		
6 <i>d</i>	87 Ir 223		88 Ra 226.05		89 Ac 227	90 Th 232.12	91 Pa 231	92 U 238.07	93 Np 237	94 Pu 239	95 Am 241	96 Cm 242	97 Bk 243	98 Cf 244	99 E (250)	100 Fm (254)	101 Mv (256)
7 <span>[5f]</span>																	
<span>[4f]</span>	58 Ce 140.13	59 Pr 140.92	60 Nd 144.27	61 Pm 147	62 Sm 150.43	63 Eu 152.0	64 Gd 156.9	65 Tb 159.2	66 Dy 162.46	67 Ho 164.94	68 Er 167.2	69 Tm 169.4	70 Yb 173.04	71 Lu 174.99			



terized by the occurrence of many valences and by the formation of complex ions which are highly coloured and are paramagnetic, while the ions of all previous elements absorb light only in the far ultra-violet and are diamagnetic. For this reason it has long been held that the  $3d$  orbits, which do not appear up to this point, are inserted here. Bohr drew this conclusion from an extrapolation which has since been confirmed by analysis of the  $\text{Sc}^{++}$  spectrum: if the values of the lowest  $S$ ,  $P$  and  $D$  terms for  $\text{K}$  and  $\text{Ca}^+$  are plotted as functions of the atomic number, the extrapolation to  $\text{Sc}^{++}$  shows that the largest term corresponding to the lowest energy state is a  $D$  term. Thus from  $Z = 21$  onward, the nineteenth electron is no longer placed in a  $4s$  orbit, but in a  $3d$  orbit. The unusual properties of the elements in the sequence  $\text{Sc}$  to  $\text{Ni}$  are attributable to the fact that all their electrons over 18 belong to an unclosed group. It is only when we get to  $\text{Cu}$  that we again find the essential spectroscopic features of an atom having one valence electron, although the occurrence of additional lines and the chemical bivalence indicate that the underlying shell is still vulnerable to outside disturbances. Since  $\text{Cu}$  has 29 electrons and has one  $4s$  electron, as has  $\text{K}$  ( $Z = 19$ ), a total of 10  $3d$  electrons have been added in the series  $\text{Sc}$  to  $\text{Ni}$ . This subsequent filling out of a provisionally completed shell repeats in the following periods. The group of four-quantum orbits is completed with  $\text{Kr}$ , but in analogy with the other rare gases, this element has only  $4s$  and  $4p$  orbits. Nevertheless, the construction of the five-quantum electron group already begins at  $\text{Rb}$ ; the filling in of this group is interrupted at  $\text{Y}$ , i.e. at the place corresponding to  $\text{Sc}$ . Again it is  $d$  orbits which are supplied to the four-quantum group in the series  $\text{Y}$  to  $\text{Pd}$ . It must be remembered, however, that the  $4f$  orbits, which are also possible, have not yet appeared.  $\text{Ag}$ , which is definitely univalent, possesses a  $5s$  electron, as does  $\text{Rb}$ . The  $5s$  and  $5p$  orbits are again closed at the element  $\text{X}$ , and the valence electron of  $\text{Cs}$  travels in a  $6s$  orbit. The filling in of the  $5d$  group is resumed with  $\text{La}$ , the third element of the series. At this point there begins a series of elements which are chemically very similar and which have no counterparts in the elements above them. These are the rare earths. There is every reason to assume that the insertion of the  $4f$  group—which is not yet present—begins here. The number of  $4f$  electrons to be supplied can be guessed by extrapolation: According to what we have seen, the relation between the quantum number and the number of electrons in a completed shell is given by

Principal quantum number $n$ ..	1	2	3
Number of electrons .. ..	$2 = 2 \cdot 1^2$	$8 = 2 \cdot 2^2$	$18 = 2 \cdot 3^2$

It is to be expected, then, that there will be in all  $2 \cdot 4^2 = 32$  four-

# ELECTRON CONFIGURATIONS OF THE ELEMENTS

		1s								
			4s	4p	4d	4f	5s	5p	5d	6s
1 H	1									
2 He	2									
		2s	2p							
3 Li	1									
4 Be	2									
5 B	2	1								
6 C	2	2								
7 N	2	3								
8 O	2	4								
9 F	2	5								
10 Ne	2	6								
		3s	3p	3d	4s					
11 Na	1									
12 Mg	2									
13 Al	2	1								
14 Si	2	2								
15 P	2	3								
16 S	2	4								
17 Cl	2	5								
18 A	2	6								
19 K	2	6			1					
20 Ca	2	6			2					
21 Sc	2	6	1		2					
22 Ti	2	6	2		2					
23 V	2	6	3		2					
24 Cr	2	6	5		1					
25 Mn	2	6	5		2					
26 Fe	2	6	6		2					
27 Co	2	6	7		2					
28 Ni	2	6	8		2					
29 Cu	2	6	10		1					
		4s	4p	4d	5s					
30 Zn	2									
31 Ga	2	1								
32 Ge	2	2								
33 As	2	3								
34 Se	2	4								
35 Br	2	5								
36 Kr	2	6								
37 Rb	2	6			1					
38 Sr	2	6			2					
39 Y	2	6	1		2					
40 Zr	2	6	2		2					
41 Nb	2	6	4		1					
42 Mo	2	6	5		1					
43 Tc	2	6	(6)	(2)						
44 Ru	2	6	7		1					
45 Rh	2	6	8		1					
46 Pd	2	6	10							
47 Ag	2	6	10		1					
48 Cd	2	6	10		2					
		5s	5p	5d	6s	6p	6d	7s	5f	
49 In	2	6	10		2	1				
50 Sn	2	6	10		2	2				
51 Sb	2	6	10		2	3				
52 Te	2	6	10		2	4				
53 I	2	6	10		2	5				
54 Xe	2	6	10		2	6				
55 Cs	2	6	10		2	6			1	
56 Ba	2	6	10		2	6			2	
57 La	2	6	10		2	6	1		2	
58 Ce	2	6	10	2	2	6	1		2	
59 Pr	2	6	10	3	2	6	(1)	(2)		
60 Nd	2	6	10	4	2	6	0		2	
61 Pm	2	6	10	5	2	6	(0)	(2)		
62 Sm	2	6	10	6	2	6	0		2	
63 Eu	2	6	10	7	2	6	0		2	
64 Gd	2	6	10	7	2	6	1		2	
65 Tb	2	6	10	9	2	6	1		2	
66 Dy	2	6	10	10	2	6	(1)	(2)		
67 Ho	2	6	10	11	2	6	(1)	(2)		
68 Er	2	6	10	12	2	6	1		2	
69 Tm	2	6	10	13	2	6	0		2	
70 Yb	2	6	10	14	2	6	0		2	
71 Lu	2	6	10	14	2	6	1		2	
		5s	5p	5d	6s	6p	6d	7s	5f	
72 Hf	2	6	2	2						
73 Ta	2	6	3	2						
74 W	2	6	4	2						
75 Re	2	6	5	2						
76 Os	2	6	6	2						
77 Ir	2	6	7	2						
78 Pt	2	6	9	1						
79 Au	2	6	10	1						
80 Hg	2	6	10	2						
81 Tl	2	6	10	2	1					
82 Pb	2	6	10	2	2					
83 Bi	2	6	10	2	3					
84 Po	2	6	10	2	4					
85 At	2	6	10	2	5					
86 Rn	2	6	10	2	6					
87 Fr	2	6	10	2	6			1	5f	
88 Ra	2	6	10	2	6			2		
89 Ac	2	6	10	2	6	1		2		
90 Th	2	6	10	2	6	2		2		
91 Pa	2	6	10	2	6	1	(2)	(2)	(2)	
92 U	2	6	10	2	6	1	2	(3)		
93 Np	2	6	10	2	6	(1)	(2)	(4)		
94 Pu	2	6	10	2	6	(1)	(2)	(5)		
95 Am	2	6	10	2	6	(1)	(2)	(6)		
96 Cm	2	6	10	2	6	(1)	(2)	(7)		
97 Bk	2	6	10	2	6	(1)	(2)	(8)		
98 Cf	2	6	10	2	6	(1)	(2)	(9)		
99 E	2	6	10	2	6	(1)	(2)	(10)		
100 Fm	2	6	10	2	6	(1)	(2)	(11)		
101 Mv	2	6	10	2	6	(1)	(2)	(12)		
102										

Tc Technetium  
 Pm Prometheum  
 At Astatine  
 Fr Francium  
 Np Neptunium  
 Pu Plutonium  
 Am Americium  
 Cm Curium  
 Bk Berkelium  
 Cf Californium  
 E Einsteinium  
 Fm Fermium  
 Mv Mendelevium

Closed shells are not repeated in the succeeding parts of the Table. Numbers in parentheses are conjectural.



quantum electrons, of which 18 have already been placed, leaving 14 rare earths. This result, found by means of what is in a sense numerical mystics, has proved to be of great practical importance—it led to the discovery of element 72, which was named Hafnium, after Copenhagen, where Bohr resides. While this element was formerly looked for among the rare earths, this suggests that it is homologous with Zr, and, as a matter of fact, the zirconium minerals contain hafnium in considerable quantity. As in the preceding periods, the filling in of the  $5d$  group is resumed with Hf and completed at Pt. Au, like Cs, has again a  $6s$  electron, and in the following elements as far as Rn (radium emanation) the rare gas configuration, consisting of  $s$  and  $p$  electrons, is again built up, so that the seven-quantum electrons begin with the following row. The subsequent insertion of the  $6d$  electrons is resumed with Ac. The construction of the five- and six-quantum electron shells is not completed, since the periodic table stops at element 92. The reason for this termination is undoubtedly that the heavy nuclei are unstable, since all elements from 84 onward are radioactive. The “transuranic” elements 93 to 98 have recently been produced artificially in the United States. As to the electron shells of these elements, the  $5f$  shell appears to begin with Ac, so that the atoms from this point onward form a new “rare earth” group that has been named the *actinide series*.

The number of electrons of the same quantum number  $n$  contained in a completed shell follows from a simple but far-reaching rule discovered by Pauli.\* If a certain direction is marked by means of an external magnetic field, the state of a spinning electron is determined by four quantities: major axis, eccentricity, and orientation of the plane of the orbit and orientation of the axis of spin. These are characterized by the quantum numbers  $n$ ,  $l = k - 1$ , and the projections  $m_l$  and  $m_s$ . Now the Pauli principle states that each combination of the four quantum numbers can occur but once in any atom, i.e. in no atom can there be two electron orbits for which all four quantities agree. If we remember that  $m_s$  can have the values  $+1/2$  and  $-1/2$ , we can dispense with the spin number by saying that in any atom there are at most two electrons for which  $n$ ,  $l$  and  $m_l$  agree. For a quantum number  $l = 1$  the projections are  $m_l = 1, 0, -1$ ; for  $l = 2$ , the possible projections are  $2, 1, 0, -1, -2$ ; in general,  $2l + 1$  different values of  $m_l$  are possible. Hence, on account of the fact that the number of possibilities is doubled by including the spin, there are 2  $s$  electrons, 6  $p$  electrons, 10  $d$  electrons, &c. Since  $l$  assumes all values from 0 to  $n - 1$ , the number of  $n$ -quantum electrons obtained by summation over all possible values of  $l$  is given by

$$N_n + \sum_0^{n-1} 2(2l + 1) = 2n^2, \quad . . . . (76)$$

\* W. Pauli, *Zeitschr. f. Phys.*, 81, p. 765 (1925).



i.e. exactly the relation given by what we referred to above as numerical mysticism!

Despite the fact that the Pauli principle offers a simple formula for the number of electrons in a completed shell, the problem of the periodic system cannot be considered fully solved until it can be shown why the construction of the three-quantum orbits, for example, is not completed in one step but is resumed at  $Z = 21$ . Attempts to solve this as a problem in atomic mechanics would lead to the hopeless task of computing the energy of each configuration. Fermi made a start toward a solution by treating the electrons in the shells as an electron gas. They are then subject to the Fermi-Dirac statistics (p. 631), and in this way it was found possible to give a very satisfactory answer to the above question.

## 12. Elements of the Theory of Band Spectra.

Spectroscopists have distinguished for a long time between *line spectra*, which are due to free atoms, and *band spectra*, in which a multitude of lines—sometimes unresolved—present the appearance of a grooved or fluted surface. Even before the development of the theory it was surmised that band spectra are to be attributed to molecules. The theory of these extremely diversified spectra is exceptionally complicated and extensive, but offers an unexpected insight into the structure of molecules. However, only an outline of the theory can be reproduced here.

The complex nature of band spectra is due to the fact that a molecule can contain and lose or gain energy not only in the form of electronic energy but also as energy of rotation and internal vibration. We confine our attention to the theory of spectra of diatomic molecules; among emission spectra these occur most frequently. Let us represent a diatomic molecule by a dumb-bell model. In addition to the energy of the electrons, the energy of this system can consist of rotation about an axis normal to the *figure axis* and oscillation of both nuclei along the figure axis.

Let us first consider pure rotation spectra, i.e. spectra resulting when only the state of rotation of the molecule changes. The quantum states of a rotator of moment of inertia  $I$  are obtained as follows. Since there is but one moment of inertia  $I$  normal to the figure axis,

$$E = E_{\text{kin}} = \frac{1}{2} I \dot{\phi}^2, \quad . \quad . \quad . \quad . \quad . \quad (77)$$

$$\text{i.e.} \quad p_{\phi} = I \dot{\phi}. \quad . \quad . \quad . \quad . \quad . \quad (78)$$

If  $\phi$  is constant, the quantum condition is

$$\oint p_{\phi} d\phi = 2\pi \phi I = m h, \quad \text{i.e.} \quad \phi = \frac{m h}{2\pi I}. \quad . \quad . \quad (79)$$

Hence the energy is 
$$E_m = \frac{m^2 h^2}{8\pi^2 I} \cdot \cdot \cdot \cdot \cdot \cdot (80)$$

On the classical theory, the rotation of a molecule can result in radiation only if the molecule is polar, i.e. when the centres of gravity of the positive and negative charges do not coincide, in which case the molecule has a positive and a negative pole. In this case the system represents a combination of two Hertzian oscillators, for which the results of p. 341 hold. This is the case, for example, for the halogen hydrides, but not for the symmetric molecules  $N_2$ ,  $O_2$ , &c. According to the Correspondence Principle, then, pure rotation band spectra are possible only with polar molecules. It follows, further, from the uniform simple rotation (absence of overtones) that  $m$  can change only by one unit, so that the transition  $m + 1 \rightarrow m$  takes place in emission. For the emitted frequency this gives

$$\nu = \frac{h}{8\pi^2 I} [(m + 1)^2 - m^2] = \frac{h}{4\pi^2 I} (m + \frac{1}{2}), \quad \cdot \quad (81)$$

i.e. there is a sequence of lines which are equidistant on the frequency scale,\* and whose separation is

$$\Delta\nu = \frac{h}{4\pi^2 I} \cdot \cdot \cdot \cdot \cdot \cdot (82)$$

Substituting this value in (81), we obtain a relation between the frequencies and their differences which contains only observable quantities:

$$\nu = \Delta\nu (m + \frac{1}{2}). \quad \cdot \cdot \cdot \cdot \cdot \cdot (83)$$

Remarkably enough, experience shows that instead of this relation, we must write

$$\nu = \Delta\nu \cdot m. \quad \cdot \cdot \cdot \cdot \cdot \cdot (83')$$

This form would be obtained by giving  $m$  half-integral values, but no reason for doing so is indicated. The same result is obtained by replacing  $m^2$  in (81) by  $m(m + 1)$ , as we did in the case of the  $g$  formula, for then we have

$$\nu = \frac{h}{8\pi^2 I} [(m + 1)(m + 2) - m(m + 1)] = \frac{h}{4\pi^2 I} (m + 1), \quad (83'')$$

which means simply an increase of the current number by one, and

\* It should be noticed that these formulæ, as they stand, are written in terms of actual frequency; it is easy to convert to wave numbers by dividing by  $c$  if desired.

thus agrees with equation (83'). If we again reduce the current number by one and once more call it  $m$ , we get the final form

$$\nu = \frac{hm}{4\pi^2 I} \cdot \cdot \cdot \cdot \cdot \cdot (83''')$$

The newer quantum theory yields precisely this expression for the energy of a rotator. Thus we see once more that the theory based on models agrees with observation only after introducing a modification.

We now go a step further and assume that the state of oscillation also experiences a change. The resulting bands are called *rotation-vibration bands*. While the pure rotation bands lie in the far infra-red—in the neighbourhood of  $30 \mu$ —the rotation-vibration bands fall in the near infra-red, in the vicinity of  $5 \mu$ . Both types of band are observed in absorption, but it is more convenient to develop the theory for emission. The absorption lines coincide with the emission lines, for a large number of molecules already possess several quanta of rotational energy on account of the thermal motion, so that the initial levels of a large number of molecules already contain several rotational quanta, in contrast with the atoms, where the thermal energy is not sufficient to provide an appreciable number of excited atoms. The absorption lines are obtained by interchanging the initial and final states.

By p. 94, the quantum states of a harmonic oscillator in which a mass  $M$  vibrates about a position of equilibrium are given by

$$E_n = h\nu_0 = \frac{h}{2\pi} \sqrt{\frac{k}{M}} \cdot \cdot \cdot \cdot \cdot (84)$$

In our model the nuclei vibrate in opposite phase, so that the centre of gravity remains at rest. The calculation of the characteristic vibration of this system is a problem quite similar to that on p. 651, where the motion of the atomic nucleus was taken into account. Here, too, we find (cf. *Ex. 43*, p. 110) that it is only necessary to replace the mass of the particle by the reduced mass  $\mu$ , where

$$\frac{1}{\mu} = \frac{1}{M_1} + \frac{1}{M_2} \cdot \cdot \cdot \cdot \cdot (85)$$

Hence the energy states of a molecule which are due to oscillation and rotation are

$$E_{n,m} = nh\nu_0 + \frac{h^2 m(m+1)}{8\pi^2 I} \cdot \cdot \cdot \cdot (86)$$

As long as the binding force on the nuclei is strictly harmonic, the Correspondence Principle again gives  $\Delta n = \pm 1$  as the selection rule for  $n$ . Nevertheless, on account of the fact that the restoring force is not perfectly harmonic, higher quantum transitions also occur; these



are referred to as overtone bands. The frequencies of the lines of the fundamental band are, according to (83'''),

$$\nu = \nu_0 \pm \frac{hm}{4\pi^2 I} \cdot \cdot \cdot \cdot \cdot \cdot (87)$$

Here the sign of the rotational part may be either positive or negative. Since the oscillational contribution is by far the greater (it displaces the system toward the near infra-red), it may happen, in emission, that a transition  $n + 1 \rightarrow n$  occurs together with a rotational change  $m - 1 \rightarrow m$  without any change in sign of the transformed energy. The lines thus appear on both sides of the null line  $\nu_0$ , which is itself absent. The quantum mechanical treatment of the oscillator also leads to half quantum number  $n$ , but this does not alter (87).

If, in addition to the energy transformations already considered, the state of the electron shells of the molecule also changes, the emitted radiation falls in the visible region or in the ultra-violet. This represents the usual band spectra which have the appearance of illuminated columns. The essentially new feature contributed by the electron transition is a change in the moment of inertia and in the interatomic bond by penetration of an electron into the electron shell which is responsible for the binding force between the atoms. We must therefore distinguish between the moment of inertia in the initial state and that in the final state, and also between frequencies of oscillation in the two states. Denoting the electronic energy by  $E_{el}$ , the energies of these two states are

$$\left. \begin{aligned} E_i &= E_{el}^{(i)} + n_i h \nu_i + \frac{h^2 m_i (m_i + 1)}{8\pi^2 I_i}, \\ E_f &= E_{el}^{(f)} + n_f h \nu_f + \frac{h^2 m_f (m_f + 1)}{8\pi^2 I_f}. \end{aligned} \right\} \cdot \cdot (88)$$

These constitute the quantum-mechanical expressions for the energy states, insofar as the electron shells possess no resultant spin. If the latter does not vanish, the *total* spin  $j$  of the molecule is an integral multiple of  $h/2\pi$  and, according to the kind of coupling of the individual spin vectors, there will be many small differences which cannot be entered into here. In order to keep the formulas manageable in such cases, we make use of the fact, mentioned above, that except for a small change in the constant term we get the same result by replacing the "quantum-mechanical square"  $x(x + 1)$  by  $(x + \frac{1}{2})^2$ . Accordingly we write

$$\left. \begin{aligned} E_i &= E_{el}^{(i)} + n_i h \nu_i + \frac{h^2}{8\pi^2 I_i} (j_i + \tfrac{1}{2})^2, \\ E_f &= E_{el}^{(f)} + n_f h \nu_f + \frac{h^2}{8\pi^2 I_f} (j_f + \tfrac{1}{2})^2. \end{aligned} \right\} \cdot \cdot (88')$$

The Correspondence Principle states that transitions to  $j_f = j$  originate in  $j_i = j + 1$  and  $j_i = j - 1$ . However, in case of a non-vanishing electron spin the transition  $j \rightarrow j$  is also permitted. If we include this case, the following formulæ result for the three branches of a band:

$$\left. \begin{aligned} R\text{- (positive) Branch: } \nu &= A + 2B(j + 1) + C(j + 1)^2, \\ P\text{- (negative) Branch: } \nu &= A - 2Bj + Cj^2, \\ Q\text{- (zero) Branch: } \nu &= A + Cj + Cj^2, \end{aligned} \right\} \quad (89)$$

where the abbreviations  $A$ ,  $B$  and  $C$  are defined as follows:

$$\left. \begin{aligned} A &= \nu_{el} + n_i \nu_i - n_f \nu_f + \frac{1}{4} \cdot \frac{h}{8\pi^2} \left( \frac{1}{I_i} - \frac{1}{I_f} \right), \\ 2B &= \frac{h}{8\pi^2} \left( \frac{1}{I_i} + \frac{1}{I_f} \right), \\ C &= \frac{h}{8\pi^2} \left( \frac{1}{I_i} - \frac{1}{I_f} \right). \end{aligned} \right\} \quad (90)$$

As mentioned, the  $Q$ -Branch is absent if the spin of the electron shells is zero. The designations  $R$ ,  $P$  and  $Q$  are arbitrary and have an historic origin.

The positions of individual lines are represented very vividly by a diagram due to Fortrat (fig. 6). In a rectangular system of coordinates take  $\nu$  as abscissa and  $m$  as ordinate. For the present assume  $m$  to be continuously variable. Then we obtain three parabolic arcs for the three branches. But only the points whose ordinates are whole numbers correspond to spectral lines. We must therefore cut these curves by the lines  $m = 1, 2, 3, \dots$  and project the points of intersection on to the  $\nu$ -axis. It is then easy to see the way the head of the band originates; it is a point toward which the lines crowd together, but is not a mathematical accumulation point like a series limit. This representation also suggests why a band has the superficial appearance of being a spatial configuration.

If we now consider the bands which belong to various values of  $n_i$  and  $n_f$ , we must remember that the harmonic binding is merely a first approximation to the power series development of the actual law of force. As a result, the selection rule for  $n$  is not obeyed. The change in the energy levels caused by the higher powers in the law of force will not be taken into account here. The contribution of the change in the state of oscillation then becomes

$$\nu_{os} = n_i \nu_i - n_f \nu_f = (n_i - n_f) \nu_i + n_f (\nu_i - \nu_f). \quad (91)$$

The second form yields a subdivision of the doubly infinite manifold if the oscillation frequency is not changed very much by the electron transition. In that case the second term is small compared with the first and we can collect all partial bands corresponding to the same change of  $n$  into one *band group*, while the entire *band system* is ob-

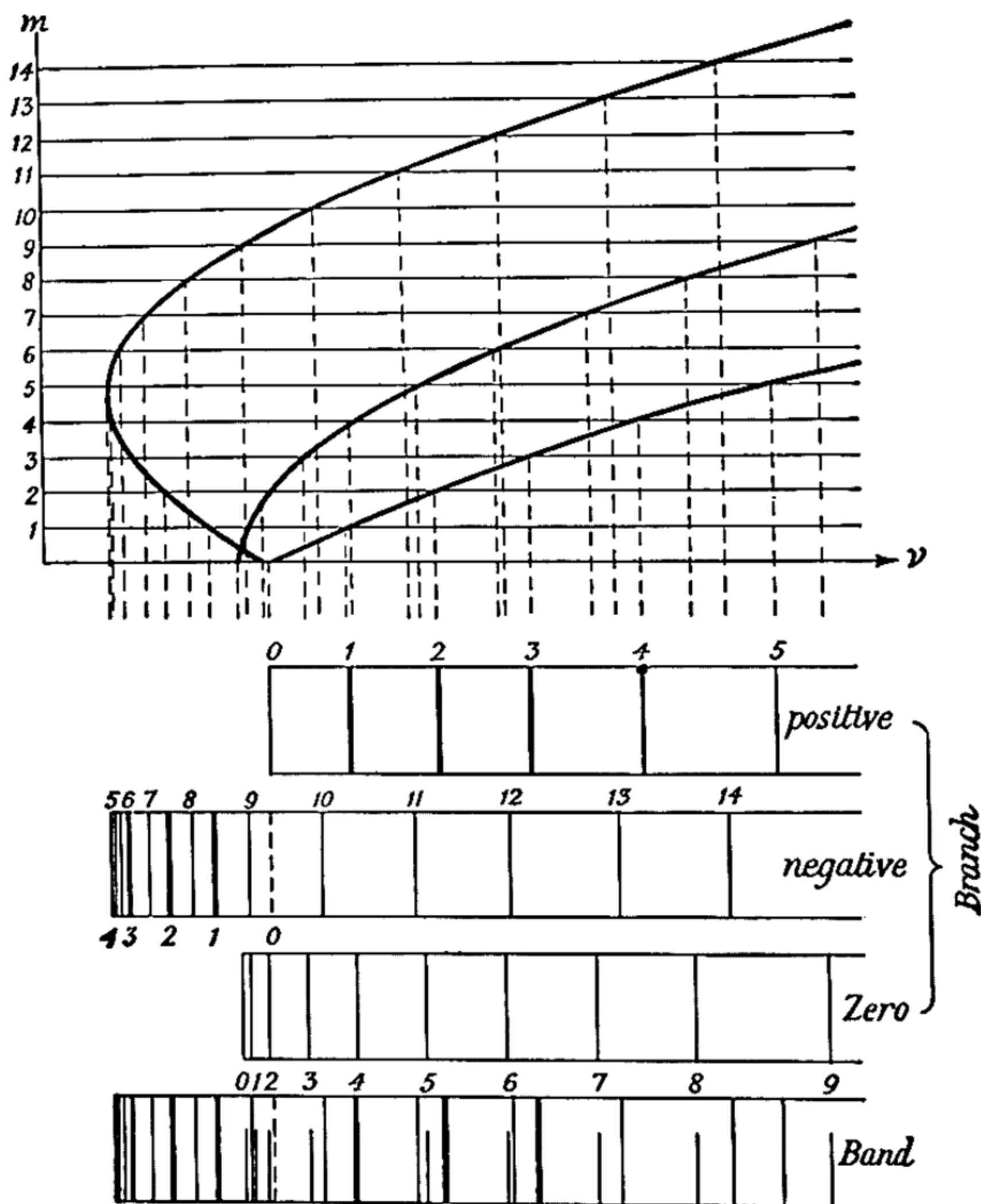


Fig. 6

tained by admitting all possible values of  $(n_i - n_f)$ . All lines belonging to the same band system correspond to the same electron transition. Thus the event which gives rise to a single line in atomic spectra is the cause of a band system consisting of many thousands of lines. However, in this connexion it must be remembered that only *one* of the many lines is emitted or absorbed as a quantum  $h\nu$  in *each* elementary process. If the electronic states of molecules were as numerous as those of atoms, we would obtain a maze of many band



systems. Fortunately the number of possible energy states of the electron shells of molecules is, evidently, considerably smaller (except for  $\text{H}_2$  and  $\text{He}_2$ ), so that a molecule generally possesses only a few band systems. The reason is that more violent disturbances of the electron shell which effects the bond lead to the decomposition of the molecule.

## CHAPTER XL

### CRITICAL ATOMIC THEORY: THE NEW QUANTUM MECHANICS

#### 1. The Wave and Particle Aspects of Light. Compton Effect and Raman Effect.

Although the phenomena of interference appear to prove that light is of the nature of waves, we have always treated emission and absorption as though the light consisted of individual energy centres (quanta) of amount  $h\nu$  which could not be subdivided. This quantum-like energy transfer has been rigorously demonstrated by an enormous number of experiments. Attempts to introduce this process into the wave picture by assuming that the atom could store up the electromagnetic energy incident on it until the quantity  $h\nu$  is reached, lead to quite impossible accumulation times for X-rays, so that this procedure was soon abandoned. Attempts to give the quanta themselves such properties that interference phenomena would result without waves proved equally unfruitful. There remains no other course but to look upon waves and quanta as two observable aspects of a single phenomenon whose true nature cannot be described in terms of any mechanical model. A crude representation of this dual property of light by assuming certain points on the wave front to be the seat of the energy also proved impossible. In no way can we spatially localize the quanta in a light wave. The true source of these difficulties will become apparent in the following section. We shall begin with the discussion of a phenomenon which more than any other brings into prominence the corpuscular nature of light, and which was therefore considered for some time to be the *experimentum crucis* for the quantum nature of light. This is the Compton effect.\*

If monochromatic X-radiation is allowed to fall on a scattering substance—e.g. a block of paraffin—and if the laterally scattered radiation is examined spectroscopically, the original line is found to have a companion on its long-wave side, the separation of the two depending upon the angle between the primary ray and the direction

\* A. H. Compton, *Phys. Rev.*, 21, p. 483 (1923); *Phil. Mag.*, 46, p. 897 (1923). The theory was given almost at the same time by P. Debye, *Physikal. Zeitschr.*, 24, p. 181 (1923).

of scattering (direction of observation). The position of the secondary line can be computed in a frankly naïve way by considering the scattering process to be the collision of a light quantum (or *photon*) with an electron. The energy delivered to the electron according to the elementary laws of impact must result in a decrease of energy—and hence a decrease in frequency—of the primary ray. We may look upon the electron as being free if it is one of the outer electrons of an atom, for in the X-ray region the quanta have energies of several tens of thousands of electron volts, while the work of removal of an outer electron is of the order of but a few electron volts. This, in fact,

is the reason that the phenomenon is observed only in the X-ray region. We now write the equations of conservation of energy and conservation of momentum for this collision process. The latter we shall decompose into its component equations for the direction of the primary ray and the direction normal to it. Denote by  $\nu'$  the fre-

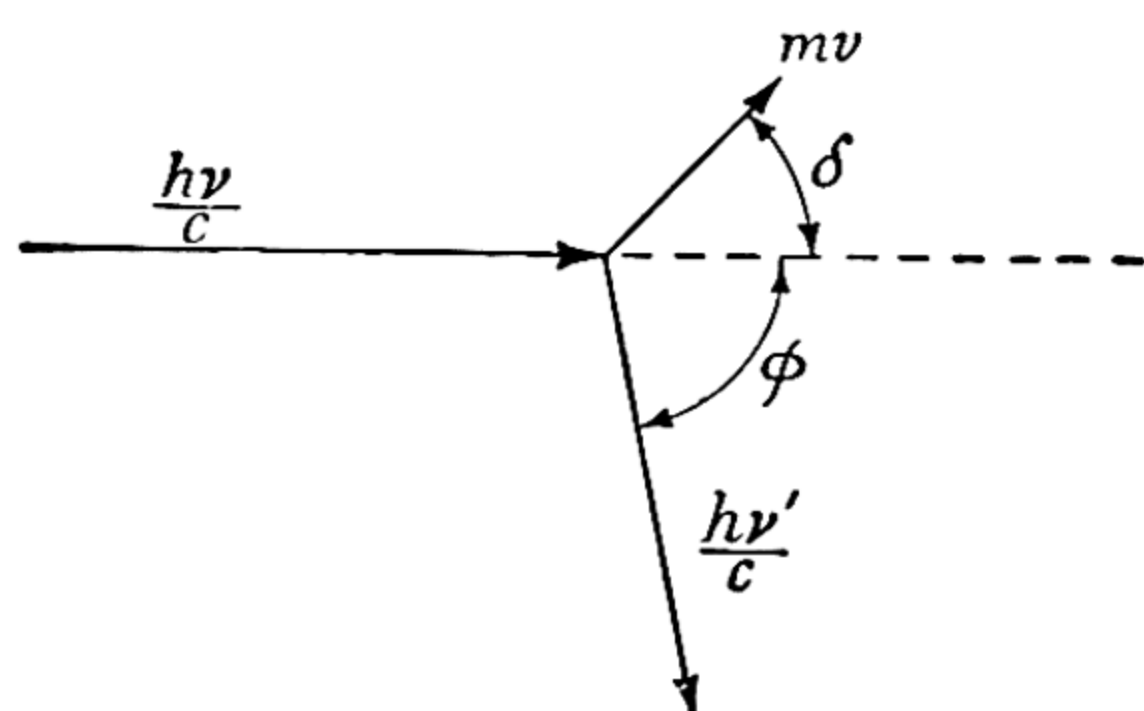


Fig. 1

quency of the scattered light, by  $m$  the mass of the electron, by  $\phi$  the angle between the primary ray and the scattered ray, and by  $\delta$  the angle between the primary ray and the velocity vector of the recoil electron (cf. fig. 1). By the principle of the inertia of energy (p. 258), the “mass” of the photon is

$$m_q = \frac{h\nu}{c^2}, \quad . . . . . (1)$$

so that its momentum is

$$m_q c = \frac{h\nu}{c}. \quad . . . . . (2)$$

We then have the energy equation

$$h\nu = h\nu' + \frac{1}{2}mv^2, \quad . . . . . (3)$$

and the two component momentum equations

$$\left. \begin{aligned} \frac{h\nu}{c} &= \frac{h\nu'}{c} \cos \phi + mv \cos \delta \\ 0 &= \frac{h\nu'}{c} \sin \phi - mv \sin \delta \end{aligned} \right\} . . . . . (4)$$

Since we are interested only in the frequency  $\nu'$  of the scattered light



as a function of the angle of scattering  $\phi$ , we eliminate  $\delta$  from the two momentum equations and obtain

$$m^2 v^2 = \frac{h^2}{c^2} (\nu^2 + \nu'^2 - 2\nu\nu' \cos \phi). \quad . . . (5)$$

For light of not too short wave-length we have  $\nu - \nu' \ll \nu$  (this approximation is not justified for  $\gamma$ -rays), so that we may replace  $\nu'$  by  $\nu$  in the term  $m\nu^2/2$  of the energy equation. From (5) we then have

$$h\nu = h\nu' + \frac{h^2 \nu^2}{mc^2} (1 - \cos \phi), \quad . . . (6)$$

whence 
$$\Delta\nu = \nu' - \nu = -\frac{2h\nu^2}{mc^2} \sin^2 \frac{\phi}{2}. \quad . . . (7)$$

This equation may be put in more elegant form by introducing the change in wave-length

$$\Delta\lambda = -\frac{c\Delta\nu}{\nu^2}$$

in place of the frequency change. Then there results

$$\Delta\lambda = \frac{2h}{mc} \sin^2 \frac{\phi}{2} = 2\lambda_0 \sin^2 \frac{\phi}{2}. \quad . . . (7')$$

The wave-length  $\lambda_0$  introduced in this way is the wave-length of radiation whose photon has an inertial mass equal to that of the electron, for then we have

$$m_a = \frac{hc}{\lambda_0 c^2} = \frac{hmc}{ch} = m. \quad . . . (8)$$

The numerical value of  $\lambda_0$  is  $24 \times 10^{-11}$  cm., so that this wave-length falls in the  $\gamma$ -ray region. Despite the fact that we neglected the relativistic variability of mass in deducing equation (7), it is strictly correct, for exactly the same final result is obtained when this effect is taken into consideration.

\ In examining the scattering of visible light by liquids or gases, Raman\* discovered an effect which is superficially similar to the Compton effect, but is fundamentally different from the theoretical standpoint. Here again displaced lines are found near the primary line, but their displacement does *not* depend on the direction of scattering, as it does in the Compton effect. This phenomenon, too, may be interpreted, in its larger features, as a division of the photon into smaller quanta in the scattering process. Part of the energy of the primary quantum is given to the scattering molecule as energy of vibration or of rotation, or else some of the molecular

\* C. V. Raman, *Ind. Jour. Phys.*, 2, March (1928). See also *Nature*, 121, p. 501 (1928); *ibid*, 122, p. 12 (1928).

energy is added to the primary photon. The equation of energy is

$$h\nu = h\nu' \pm nh\nu_0, \quad . \quad . \quad . \quad . \quad . \quad . \quad (9)$$

where  $\nu_0$  must be one of the characteristic frequencies of the molecule, as determined from its band spectra. The details of this phenomenon require a more exact theory, the basic features of which will be given in Chap. XLI, § 3 (p. 716).

## 2. The Inexactness of Atomic Observations.

Heisenberg was the first to recognize the fact that the motions of electrons in the atom are, as a matter of essential principle, incapable of being verified experimentally. In celestial mechanics we can check our calculations by computing the future configuration of the heavenly bodies from their instantaneous positions and velocities and then comparing the results with observation. For the electron system of an atom, however, the rigorous determinism of classical mechanics is lacking—we cannot ascertain the necessary initial conditions. This is inherently impossible, quite apart from technical difficulties, the reason being that any means of observation by which atomic processes are conveyed to our perceptions causes a disturbance of the electron system of the atom.

If we allow our imagination free rein and disregard all technical difficulties, the first thing we might try is to determine the instantaneous positions of the electrons with the aid of a microscope. The limit of resolution of this instrument is given by  $\lambda/2A$ , where  $A$  is the numerical aperture, which we take equal to unity—i.e. we take the angular opening of the incident pencil to be  $\pi/2$ . If we now demand that the position of the electron is to be determined to within at least 1 per cent of the radius of the atom ( $10^{-8}$  cm.), we must employ light whose wave-length is of the order of  $10^{-11}$  cm., i.e. 1 X.U. The relative change in wave-length caused by the Compton effect is extremely large for such short waves, as one sees from the fact that for the characteristic wave-length 24 X.U. and a scattering angle of  $\pi/2$  we already have  $\Delta\lambda = \lambda$ . From (5) the momentum of the recoil electron is approximately \*

$$mv = \frac{2h\nu}{c} \sin \frac{\phi}{2}. \quad . \quad . \quad . \quad . \quad . \quad (10)$$

This is the change in momentum of the electron resulting from its observation by means of  $\gamma$ -rays. Since we assume a range in angle of from 0 to  $\pi/2$ , corresponding to the aperture of the microscope, the momentum is determined only to the quantity

$$\Delta p = mv \approx \frac{h}{\lambda}, \quad . \quad . \quad . \quad . \quad . \quad (11)$$

\* Putting  $\nu' \approx \nu$ .



since all the light contributing to the image must be scattered by the object and so must participate in the Compton effect. We may measure the velocity by analysing the scattered light spectrally, determining the velocity of the scattering particle from the change in wave-length caused by the Doppler effect. But we do not know how to take into account the velocity change due to the Compton effect. The original velocity might have been changed by the act of observation, in which case the error of observation lies in the range  $\Delta p$ . If we use longer waves in order to reduce the Compton effect we make the determination of position uncertain. On account of diffraction, the error  $\Delta q$  in the position determination is of the order of magnitude of  $\lambda$ , hence the product of the error in the position co-ordinate by that in the momentum co-ordinate is

$$\Delta p \Delta q \approx h. \quad . \quad . \quad . \quad . \quad . \quad . \quad (12)$$

This equation represents one of the most important principles in theoretical physics; it shows us the limits of applicability of mechanical causality. It has pointed out new paths in atomic physics: no longer is it a task of the theory to disclose subatomic processes; rather, the theory must—without considering phenomena which are inaccessible to observation—create a mechanics which reproduces the observable quantities—the spectral lines and the energy levels—from a model, of which we are required to know only the component parts and the acting forces but not the mechanical constants of integration. From this point of view we can understand, for example, why attempts to explain chemical bonds by introducing phase relationships between the electron orbits of the constituents were unsuccessful.

The indeterminacy relation, which holds for any pair of canonically conjugate variables, becomes of especial importance when applied to energy and time:\*

$$\Delta E \cdot \Delta t \approx h. \quad . \quad . \quad . \quad . \quad . \quad . \quad (13)$$

Since the quantum energy states are usually thought of as being absolutely sharp, this implies that we can say nothing regarding the time an atom remains in such a state. Actually, it is found that after eliminating all broadening influences like the Doppler effect, a spectral line still has a finite width, so that the energy levels must be considered to have a finite—although very small—breadth. From this we deduce a time interval of the order of  $10^{-8}$  sec., which is interpreted as the time the state persists.

These considerations throw new light on the failure of all attempts to treat theoretically the process of emission itself on the basis of the concepts developed in the preceding chapter. Our ordinary

\* In order to see that these quantities constitute a pair of canonically conjugate variables, consider the second of Hamilton's Equations on p. 122. If we put  $H = E = p$  it follows that  $dq/dt = 1$ , i.e.  $q = t$ .



ideas would lead us to require of such a theory that it should furnish an equation which gives the amount of energy radiated at each instant, i.e.  $E = f(t)$ . However, such a relation could not be tested experimentally. If, for example, we divide the time of emission, which is empirically known to be of the order of  $10^{-8}$  sec., into  $10^7$  intervals—a continuous function must permit of any arbitrarily small division of the variables—and if we require that it be possible to give the energy content of the atom for each of these intervals, we find that we are asking the impossible, for the indeterminacy relation gives for the error in the energy

$$\Delta E = \frac{6.6 \times 10^{-27}}{10^{-15}} = 6.6 \times 10^{-12} \text{ erg.}$$

But the work of excitation of, say, the line  $\lambda 2537$  is about 5 electron volts, or

$$E = \frac{5 \times 4.8 \times 10^{-10}}{300} = 8 \times 10^{-12} \text{ erg.}$$

Comparison of the results shows that the error in the energy is about as large as the entire amount of energy available. Parenthetically, it might be remarked that for the radiation from a macroscopic antenna, on the other hand, neither the error in the time nor that in the energy is important, on account of the magnitude of these quantities, so that the description of the process of emission by means of a Hertzian oscillator adequately represents the facts.

If we extend the indeterminacy relation, which was originally proved for electrons, to all the elementary particles—electrons, nuclei, photons—for which analogous hypothetical experiments can be devised, we see at once why it is experimentally impossible to localize the quantum in a monochromatic wave train of frequency  $\nu$ . In an infinitely long wave train—in this case only is the wave monochromatic—the momentum  $h\nu/c$  of the photons can be given exactly by means of spectroscopic determination of the wave-length, but this makes the position co-ordinates completely indeterminate. On the other hand, if the time at which a photon passes a given point in space is confined between certain limits by using a wave train of finite length, the light is then no longer monochromatic, for it can be represented by a band of infinitely long trains of Fourier waves (cf. p. 57). In the curve which gives the intensity as a function of the frequency, the distance between the two frequencies for which the intensity has half its maximum value is given by

$$\Delta\nu = \frac{1}{\Delta t} \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad (14)$$

If we now use the  $h\nu$  law to connect the frequency with the energy

$E$  of the photons we obtain the indeterminacy relation (13) from the equation (14) which is derived solely from considerations of wave kinematics. Thus the light waves accomplish for the photons exactly what we need to confirm the Heisenberg indeterminacy. It then becomes quite natural to try to find a similar correlation of waves with the motion of electrons. The idea of "matter waves" was first propounded by L. de Broglie. This will be considered in detail in the following sections.

### 3. Matter Waves.

For light, the paths of the photons—the rays—are the orthogonal trajectories of the wave surfaces. As long as all apertures and screens are large compared with the wave-length of the light we can neglect diffraction phenomena and can compute the paths of the rays by means of Fermat's Principle

$$\delta \int_{P_1}^{P_2} n ds = \delta \int_{P_1}^{P_2} \frac{ds}{u} = 0 \quad . . . . (15)$$

( $u$  is the phase velocity) if we have given the index of refraction as a function of position. But as soon as we pass to objects whose size is of the order of magnitude of a light wave, simple geometrical optics is no longer applicable and the phenomena connected with diffraction can no longer be described in terms of light rays. We must in fact consider the wave disturbance which exists at a given surface and compute the subsequent spread of this disturbance (cf. p. 376).

On the other hand, a similar extremal principle—that of Hamilton (p. 123)—holds for the paths of particles which obey the classical mechanics. The analogy with optics becomes particularly evident if we restrict the comparison orbits to those of the same total energy. This is equivalent to passing from Hamilton's Principle to that of Maupertuis. By Hamilton's Principle we have for a particle,

$$0 = \delta \int_{t_0}^{t_1} (2T - E) dt = \delta \int_{t_0}^{t_1} 2T dt = \delta \int_{t_0}^{t_1} mv^2 dt.$$

We can now introduce the length of path as the variable of integration in place of the time:

$$0 = \delta \int_{t_0}^{t_1} mv \frac{ds}{dt} dt = \delta \int_{P_1}^{P_2} mv ds,$$

or since  $m$  is constant,

$$\delta \int_{P_1}^{P_2} v ds = 0. \quad . . . . (16)$$

Comparing (15) and (16), we see that if we wish to associate waves with particles moving according to the laws of mechanics in the same

way as light waves are associated with light rays, we must set the phase velocity  $u$  proportional to  $1/\nu$  in order that Fermat's Principle shall be valid for the matter waves. Again, if we denote the potential energy by  $U$ ,

$$v = \sqrt{\frac{2}{m}(E - U)}, \quad \text{i.e.} \quad u = \frac{C}{\sqrt{E - U}}. \quad \cdot \cdot \cdot (17)$$

If we speak of a wave function we must introduce a frequency. It is natural to do this by means of the old  $h\nu$  relation:

$$E = h\nu, \quad u = \frac{C}{\sqrt{h\nu - U}}. \quad \cdot \cdot \cdot (18)$$

This means that the phase velocity depends upon  $\nu$ ; hence there exists dispersion of our tentatively introduced waves.

Now in a dispersive medium it is necessary to distinguish between the phase velocity  $u$  and the group velocity  $u_g$ . The latter is the velocity which we can measure directly by means of a recording mechanism, while the individual extrema in a wave train of constant amplitude are not distinguishable. For a material particle, on the other hand, we can, of course, measure the actual velocity. Hence we demand that the group velocity of the waves associated with a particle shall coincide with the velocity of the particle. Using the relation between group velocity and phase velocity given on p. 66, and making  $u_g = v$ , we have

$$\frac{1}{u_g} = \frac{d\left(\frac{\nu}{u}\right)}{d\nu} = \frac{1}{v} = \frac{1}{\sqrt{\frac{2}{m}(E - U)}}. \quad \cdot \cdot \cdot (19)$$

Employing the  $h\nu$  law and integrating,

$$\frac{\nu}{u} = \int \frac{d\nu}{\sqrt{\frac{2}{m}(h\nu - U)}} = \frac{1}{h} \sqrt{2m(h\nu - U)}. \quad \cdot (20)$$

We can therefore state a dispersion law which fulfils the requirements  $1/u \propto \nu$  and  $u_g = v$  if we replace the constant  $C$  in equation (18) by  $h\nu/\sqrt{(2m)}$ .\*

We then have

$$u = \frac{h\nu}{\sqrt{2m(h\nu - U)}}. \quad \cdot \cdot \cdot (21)$$

\* In equation (18),  $C$  can still contain  $E$  and hence  $h\nu$ , since all comparison paths are assumed to have the same energy.



The wave-length of the matter waves is given by

$$\lambda = \frac{u}{\nu} = \frac{h}{\sqrt{2m(h\nu - U)}} = \frac{h}{\sqrt{2m(E - U)}} \quad \cdot \cdot \cdot (22)$$

In a region free of force,

$$\lambda_0 = \frac{h}{\sqrt{2mE}} = \frac{h}{mv} \quad \cdot \cdot \cdot \cdot \cdot (22')$$

Passing from a region where  $U = 0$  to one where the potential  $U$  exists corresponds to entering a region whose index of refraction is

$$n = \sqrt{\frac{E - U}{E}} \quad \cdot \cdot \cdot \cdot \cdot (23)$$

If  $U$  is a function of position, we have a medium of variable index of refraction in which the orthogonal trajectories of the wave surfaces—i.e. the orbits—are curved.

If, for example, we take  $U$  to be the potential of gravity and construct, according to geometrical optics, the rays of equal  $E$  passing through a fixed point, we obtain all parabolic trajectories having the same initial velocity.

Up to this point the analogy between the mechanics of a particle and geometrical optics may seem to be mere idle speculation. If it is to be more than that, it must be possible to detect phenomena which justify the introduction of associated wave fields—i.e. interference and diffraction phenomena—for particles, as was done in optics to prove the wave nature of light. If the “de Broglie wave-length” of an electron which has traversed a difference of potential  $V$  is calculated by (22'), we obtain the simple relation

$$\lambda_0 = \sqrt{\frac{150}{V}}, \quad \cdot \cdot \cdot \cdot \cdot (24)$$

where  $V$  is measured in volts and  $\lambda$  is given in A.U. This shows that electrons whose velocity is several hundred or several thousand volts are associated with wave-lengths corresponding to those of X-rays. Hence we may expect the same diffraction and interference phenomena as for X-rays; in fact, Davisson, Kunsman and Germer had previously found selective maxima when a beam of electrons was reflected from the surface of a nickel crystal at given glancing angles corresponding to those of X-rays of the same wave-length.\* The detection of the “matter waves” is even simpler by the Debye-Scherrer method. It is necessary only to send a narrow ray of electrons of uniform velocity through a thin metal foil; diffraction by the

\* See C. Davisson and L. Germer, *Nature*, 119, p. 558 (1927); *Phys. Rev.*, 30, p. 705 (1927).

variously oriented crystals gives a ring diagram which is as perfect as any obtained by means of X-rays. At those places where the wave computation indicates destructive interference no electrons are found, while at maxima we observe a large number of electrons.

Stern has even observed matter wave interference by reflecting a beam of helium atoms from the surface of a crystal.\* According to equation (22'), the greater mass in this case yields a wave-length 7300 times as short, for the same velocity. But this is compensated by the fact that the velocity (of thermal agitation) is much smaller in this instance, so that the de Broglie wave-lengths again fall in the region which is readily measured by means of crystals.

If we wish to account for the more detailed phenomena of diffraction in optics, we must turn to the wave *differential* equation. Similarly, in atomic mechanics, where we also deal with quantities comparable with the wave-length, we must use an analogous wave equation

$$\Delta\Psi = \frac{1}{u^2} \frac{\partial^2\Psi}{\partial t^2}, \quad . . . . . (25)$$

where

$$u = \frac{h\nu}{\sqrt{2m(h\nu - U)}} = \frac{E}{\sqrt{2m(E - U)}}. \quad . . . (26)$$

The meaning of the quantity  $\Psi$  which satisfies the wave equation is still undetermined. We require of  $\Psi$  only that its square—or, if it is complex, its norm  $\Psi\bar{\Psi}$ —be a measure of the intensity of, for example, an electron beam—just as in any optical theory the square of the quantity which satisfies the wave equation is a measure of the light intensity, without necessarily requiring that it be a component of the electric vector.

As is always done in the wave theory, we take a simple periodic function to represent the dependence on time:

$$\Psi = \psi e^{2\pi i \nu t} = \psi e^{2\pi i \cdot E/h \cdot t}, \quad . . . . . (27)$$

and obtain for the variation of  $\psi$  with position the fundamental Schrödinger equation †

$$\Delta\psi + \frac{8\pi^2 m}{h^2} (E - U)\psi = 0. \quad . . . . . (28)$$

This equation has the same significance in atomic mechanics as has the Hamilton-Jacobi equation

$$H\left(\frac{\partial S}{\partial q_k}, q_k\right) = E \quad . . . . . (29)$$

in celestial mechanics.

\* See F. Knauer and O. Stern, *Zeitschr. f. Phys.*, **53**, 11–12, p. 779 (1929).

† E. Schrödinger, *Phys. Rev.*, **28**, p. 1049 (1926); *Collected Papers on Wave Mechanics*, pp. 2, 77.

#### 4. The Proper Values of the Wave Equation.

The wave function  $\Psi$  governs the motion of the electrons in the same way as light waves determine the motion of the photons—wherever  $\Psi$  vanishes there are no electrons. If, then, we seek a solution which is to represent the motion of an electron in an atom,  $\Psi$  must certainly vanish at infinity, for the bound electrons are all confined to the finite region. In addition, we must demand that any physically useful solution is to be a continuous and a single-valued function of position in the finite region. Simple as these conditions seem, they are, nevertheless, sufficient to determine the characteristic values which in the old quantum mechanics were singled out from among the many mechanically possible solutions only by imposing an auxiliary condition. It may be shown that the condition of single-valuedness is identical with the old quantum conditions for a sufficiently large periodic orbit. To demonstrate this we shall confine our attention to systems with one degree of freedom. As a matter of fact, the extension to more degrees of freedom—and hence, in general, to multiply-periodic systems—does not present the slightest difficulty. Let us consider a closed orbit to be the orthogonal trajectory of wave surfaces. Then if we move along an orbit,  $\Psi$  must return to its former value when we again reach the starting-point, in harmony with our requirement that  $\Psi$  is to be single valued. This means that the phase of the wave motion must have increased by an integral multiple of  $2\pi$ , and this in turn implies that a whole number of wave-lengths must be contained in the path. In general, the wave-length may vary from place to place if the index of refraction is variable. We may therefore express the condition of single-valuedness as follows:

$$\oint \frac{dq}{\lambda} = \oint \frac{1}{h} \sqrt{2m(E - U)} dq = n. \quad . \quad . \quad . \quad (30)$$

But according to the Hamilton-Jacobi equation (p. 127), we have

$$\frac{1}{2m} \left( \frac{dS}{dq} \right)^2 + U = E, \quad \text{i.e.} \quad p = \frac{dS}{dq} = \sqrt{2m(E - U)}. \quad (31)$$

We thus obtain the old quantum condition

$$\oint p dq = nh. \quad . \quad . \quad . \quad . \quad . \quad (32)$$

It might seem, at first, that this merely secures a more reasonable foundation for the quantum conditions without altering the result in the least. At the same time it must be remembered that the single-value condition is given a significant meaning by equation (30) only if the change in wave-length is so gradual that the whole orbit can be divided into intervals  $dq$ , within which several equal wave-



lengths lie. Since this condition is by no means always satisfied, it is quite possible that a strict calculation based upon the wave equation may lead to somewhat different results.

The theory of linear differential equations shows that the condition of single-valuedness, or the boundary conditions, can be satisfied only for certain values of the parameter  $E$ . These are called the *proper values*, or *characteristic values*,\* of the differential equation (cf. the simple cases discussed on p. 186 *et seq.*). Our chief task, therefore, is the determination of the proper values for a given mechanical problem; these are, by their very nature, the selected energy levels. The corresponding *characteristic functions*, or *proper functions*, on the other hand, are only of secondary importance. We shall illustrate the process by two examples—the hydrogen atom and rotational band spectra.

### 5. The Hydrogen Atom.

Since the potential energy

$$U = -\frac{e^2}{r} \quad . \quad . \quad . \quad . \quad . \quad . \quad (33)$$

depends only upon  $r$ , the co-ordinates appropriate to the problem are spatial polar co-ordinates. Denoting the electron mass by  $m_0$ , the Schrödinger equation (28) (p. 696) becomes (cf. p. 268)

$$\begin{aligned} \frac{\partial^2 \psi}{\partial r^2} + \frac{2}{r} \frac{\partial \psi}{\partial r} + \frac{1}{r^2} \left\{ \frac{1}{\sin \theta} \frac{\partial}{\partial \theta} \left( \sin \theta \frac{\partial \psi}{\partial \theta} \right) + \frac{1}{\sin^2 \theta} \frac{\partial^2 \psi}{\partial \phi^2} \right\} \\ + \frac{8\pi^2 m_0}{h^2} \left( E + \frac{e^2}{r} \right) \psi = 0. \quad . \quad (34) \end{aligned}$$

We try to solve this equation by a value of  $\psi$  which is the product of separate functions of  $r$ ,  $\theta$  and  $\phi$ :

$$\psi = R(r) \Theta(\theta) \Phi(\phi). \quad . \quad . \quad . \quad . \quad (35)$$

So far as  $\phi$  is concerned, we can have only the function

$$\Phi = e^{\pm im\phi}, \quad . \quad . \quad . \quad . \quad . \quad (36)$$

or a linear function of  $\cos m\phi$  and  $\sin m\phi$ , with integral values of  $m$  on account of the single-value condition. Whatever function we take, we have always

$$\frac{\partial^2 \psi}{\partial \phi^2} = -m^2 R \Theta \Phi. \quad . \quad . \quad . \quad . \quad (37)$$

\* Ger. *Eigenwerte*, *Eigenfunktionen*; Fr., *constantes fondamentales*, *fonctions fondamentales*.

Eliminating  $\Phi$ , we obtain the following equation which depends only upon  $r$  and  $\theta$ :

$$\frac{r^2}{R} \left\{ \frac{d^2 R}{dr^2} + \frac{2}{r} \frac{dR}{dr} + \frac{8\pi^2 m_0}{h^2} \left( E + \frac{e^2}{r} \right) R \right\} \\ = - \frac{1}{\Theta} \left\{ \frac{1}{\sin \theta} \frac{d}{d\theta} \left( \sin \theta \frac{d\Theta}{d\theta} \right) - \frac{m^2 \Theta}{\sin^2 \theta} \right\}, \quad . . . \quad (38)$$

which we have written in such a way that the left member contains only functions of  $r$ , the right member only functions of  $\theta$ , so that each member must be a constant, say  $\alpha$ , so that the variables are separated. First we examine the equation involving  $\theta$ :

$$\frac{1}{\sin \theta} \frac{d}{d\theta} \left( \sin \theta \frac{d\Theta}{d\theta} \right) + \left( \alpha - \frac{m^2}{\sin^2 \theta} \right) \Theta = 0. \quad . \quad (39)$$

This equation is familiar as the differential equation of spherical harmonics. For details, see the Mathematical Addendum, § 4 (p. 819). We substitute

$$\cos \theta = x, \quad . . . . . \quad (40)$$

and write in place of  $\Theta(\cos \theta)$  the more common symbol  $y(x)$ ; there results

$$(1 - x^2) \frac{d^2 y}{dx^2} - 2x \frac{dy}{dx} + \left( \alpha - \frac{m^2}{1 - x^2} \right) y = 0. \quad . \quad (41)$$

As is shown in the Theory of Spherical Harmonics, this equation has solutions which satisfy the conditions of single-valuedness and continuity if  $\alpha$  is of the form

$$\alpha = (k + m)(k + m + 1), \quad . . . . . \quad (42)$$

where  $k$  and  $m$  are whole numbers. We set  $k + m = l$  and call the solutions corresponding to  $\alpha = l(l + 1)$  spherical harmonics of  $l$ th order. There are  $l + 1$  different functions of this kind, since  $m$  can assume all values from 0 to  $l$ . For  $m = 0$  we obtain the so-called *surface zonal harmonics* or *Legendre polynomials*. These are simple polynomials of degree  $l$ , and are most readily obtained as the coefficients of the powers of  $r$  in the expansion of the expression

$$F(r, x) = (1 - 2rx + r^2)^{-1/2}.$$

This gives

$$\left. \begin{aligned} P_0 &= 1 & P_3 &= \frac{5}{2} x^3 - \frac{3}{2} x \\ P_1 &= x & P_4 &= \frac{35}{8} x^4 - \frac{15}{4} x^2 + \frac{3}{8} \\ P_2 &= \frac{3}{2} x^2 - \frac{1}{2} & P_5 &= \frac{63}{8} x^5 - \frac{35}{4} x^3 + \frac{15}{8} x, \text{ \&c.} \end{aligned} \right\} \quad (43)$$

As one may easily verify, the polynomial of degree  $l$  is a solution of that differential equation for which the value of  $\alpha$  is  $l(l+1)$ ; for example,  $P_5$  corresponds to  $\alpha = 30$ . When the value of  $m$  is different from zero we again obtain polynomials as solutions. These are called the *associated* harmonics, and are related to the surface zonal harmonics by the equation

$$P_l^m(x) = (1-x^2)^{m/2} \frac{d^m P_l}{dx^m}. \quad \dots \quad (44)$$

Since the  $l$ th derivative of  $P_l$  is a constant, the spherical harmonic belonging to  $m = l$ , which corresponds to the highest azimuthal subdivision, is

$$P_l^l = K(1-x^2)^{l/2} = K(\sin \theta)^l. \quad \dots \quad (45)$$

We have still to determine  $R$  as a function of the radius  $r$ . If we substitute the value  $\alpha = l(l+1)$  in the separated equation (38) (p. 699), the differential equation for  $R$  becomes

$$\frac{d^2 R}{dr^2} + \frac{2}{r} \frac{dR}{dr} + \left( \frac{8\pi^2 m_0 E}{h^2} + \frac{8\pi^2 m_0 e^2}{h^2 r} - \frac{l(l+1)}{r^2} \right) R = 0. \quad (46)$$

We introduce the abbreviations

$$\left. \begin{aligned} \frac{8\pi^2 m_0 E}{h^2} &= A, & -l(l+1) &= C, \\ \frac{4\pi^2 m_0 e^2}{h^2} &= B \end{aligned} \right\}, \quad \dots \quad (47)$$

and obtain 
$$\frac{d^2 R}{dr^2} + \frac{2}{r} \frac{dR}{dr} + \left( A + \frac{2B}{r} + \frac{C}{r^2} \right) R = 0. \quad \dots \quad (48)$$

The nature of the solutions depends essentially on the sign of  $A$ , and hence of  $E$ , corresponding to the case of an elliptic or a hyperbolic orbit, as may be seen from the behaviour of the equation for large values of  $r$ . In this case the terms not containing  $1/r$  dominate, so that the equation may be replaced approximately by

$$\frac{d^2 R}{dr^2} + AR = 0. \quad \dots \quad (49)$$

According to the sign of  $A$  we have either an oscillating function as a solution or one which tends to zero or infinity exponentially.

Consider first the case when  $A$ , and therefore  $E$ , is negative. It turns out that this corresponds to the case of an elliptic orbit. Set

$$A = -\frac{1}{r_0^2}, \quad \dots \quad (50)$$





the *empirical* rule of the old theory according to which it was necessary to count the number  $l$  from 0 to  $n - 1$  instead of from 1 to  $n$ . In the same way, the  $(2l + 1)$  different values of  $m_l$  which we interpreted as the number of integral projections of  $l$ , are here found to be the number of surface harmonics  $P_l^m(x)e^{\pm im\phi}$  which, by equation (44) (p. 700) correspond to a given numerical value of  $l$ . The factor 2 results from the double sign of the exponent which, however, gives but one value for  $m = 0$ . The proper function corresponding to the proper value  $E_n$  is

$$\psi_n = e^{-\rho/2} \rho^l w_{n-l-1}(\rho) P_l^m(\cos \theta) e^{\pm im\phi}, \quad . \quad . \quad . \quad (60)$$

where the degree  $p$  of the polynomial  $w_p$  is expressed in terms of  $n$  and  $l$ . The factor  $e^{-\rho/2}$  guarantees the vanishing of the  $\psi$  function at infinity. Since we wish to compute with real  $\psi$  functions, we have not put  $e^{\pm im\phi}$  for  $\Phi$ . A second solution, corresponding to the same proper value, is the same expression with  $\sin m\phi$  replacing  $\cos m\phi$ .

There are  $n^2$  different proper functions corresponding to each definite proper value  $E_n$ , for  $l$  runs from 0 to  $n - 1$  for a given value of  $n$ , and to every value of  $l$  there are, by the above,  $2l + 1$  surface harmonics, so that the number of possible proper functions is given by

$$z_n = \sum_0^{n-1} (2l + 1) = \frac{n}{2} (1 + 2n - 2 + 1) = n^2. \quad . \quad (61)$$

The case where several proper functions correspond to one and the same proper value is what we refer to as a case of *degeneracy* in celestial mechanics, since the number of independent frequencies is smaller than the number of degrees of freedom. For this reason we also refer to such cases in wave mechanics as degenerate. The degree of degeneracy, however, is different here—if  $z$  proper functions are associated with one proper value, we say the degree of degeneracy of the system is  $(z - 1)$ .

With regard to positive values of  $E$ , which correspond to hyperbolic orbits, we may mention without proof that no discrete proper values exist in this case—every positive value of  $E$  is compatible with our requirement that the solution should be one-valued. This corresponds to the continuous spectra found at the series limits; formerly their existence was explained in an analogous way by pointing out that the kinetic energy with which an electron departs from the atom is not subject to a quantum condition.

## 6. Rotation Spectra of Diatomic Molecules.

For simplicity we shall consider a diatomic molecule, one of whose atoms is of infinite mass, the other being of mass  $M$  and at a distance

$a$  from the first. The co-ordinates appropriate to the problem are again spatial polar co-ordinates. We write

$$a^2 M = I, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (62)$$

and since there is otherwise no dependence on  $r$  and no potential energy, the wave equation simplifies to

$$\frac{1}{\sin \theta} \frac{\partial}{\partial \theta} \left( \sin \theta \frac{\partial \psi}{\partial \theta} \right) + \frac{1}{\sin^2 \theta} \frac{\partial^2 \psi}{\partial \phi^2} + \frac{8\pi^2 I E}{h^2} \psi = 0. \quad . \quad (63)$$

This equation holds quite generally if we take  $I$  to be the moment of inertia about an axis through the centre of gravity normal to the figure axis. Comparing this equation with equation (34) (p. 698), we see that we again have the equation of surface harmonics, whose proper values are given by

$$\frac{8\pi^2 I E}{h^2} = l(l+1), \quad \text{or} \quad E = \frac{l(l+1)h^2}{8\pi^2 I}. \quad . \quad . \quad (64)$$

But this is precisely the form given in § 12 of the preceding chapter (p. 682) for the rotational energy levels of diatomic molecules; the old quantum theory gives this incorrectly as  $l^2 h^2 / 8\pi^2 I$ .\* Thus the newer theory justifies the curious rule of the replacement of  $x^2$  by  $x(x+1)$ . In general, the result is that wherever the older atom mechanics yielded  $x^2$ , the square of spin quantum numbers, these are to be replaced by  $x(x+1)$ , the so-called "quantum-mechanical square".

## 7. The Physical Meaning of the $\Psi$ Function. Intensity Relations for Spectral Lines.

Schrödinger originally interpreted his  $\Psi$  function in the following way. The electron is not a localized charge within the atom, but charge and mass are "smeared" over a certain region. The density of charge is assumed to be proportional to  $\Psi \bar{\Psi}$ , the "norm", or square of the amplitude of the  $\Psi$  function. Assume first that only a single state of the atom, corresponding to the proper value  $E_m$ , is developed. The corresponding  $\Psi$  function is

$$\Psi_m = \psi_m e^{2\pi i (E_m/h)t}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (65)$$

Since  $\psi_m$  is real, we see that on multiplying by  $\bar{\Psi}_m$  the time factor disappears, so that the distribution of charge is constant in time. A configuration of this kind cannot radiate, in agreement with the first Bohr postulate.

\* Writing  $l$  instead of  $m$ .



For what follows it is necessary that the  $\psi$  functions be orthogonal, and "normalized to unity". By the latter expression we mean that, since there is an arbitrary multiplicative constant in the solution of the differential equation, we can choose it so as to make the integral of the square of any proper function  $\psi_m$ , taken over the entire region, equal to unity. We thus specify that all proper functions are normalized, i.e.

$$\int \psi_m^2 d\tau = 1. \quad . \quad . \quad . \quad . \quad . \quad . \quad (66)$$

The orthogonal property implies that the integral over the entire region of the product of two proper functions belonging to *different* proper values is zero:

$$\int \psi_m \psi_n d\tau = 0. \quad . \quad . \quad . \quad . \quad . \quad . \quad (67)$$

The proof is based upon Green's Theorem (cf. p. 270):

$$\int (\psi_m \Delta \psi_n - \psi_n \Delta \psi_m) d\tau = \oint (\psi_m \text{grad } \psi_n - \psi_n \text{grad } \psi_m) d\mathbf{S}. \quad (68)$$

The proper function  $\psi_m$  satisfies the equation

$$\Delta \psi_m + \frac{8\pi^2 m_0}{h^2} (E_m - U) \psi_m = 0, \quad . \quad . \quad . \quad (69)$$

and  $\psi_n$  satisfies the equation

$$\Delta \psi_n + \frac{8\pi^2 m_0}{h^2} (E_n - U) \psi_n = 0. \quad . \quad . \quad . \quad (70)$$

If we multiply the first equation by  $\psi_n$ , the second by  $\psi_m$ , and subtract the first from the second, we obtain

$$\psi_m \Delta \psi_n - \psi_n \Delta \psi_m = \frac{8\pi^2 m_0}{h^2} (E_m - E_n) \psi_n \psi_m. \quad . \quad (71)$$

This equation is multiplied by  $d\tau$  and integrated over all space. The integral on the left is transformed into a surface integral over an infinitely distant boundary by means of Green's Theorem (68). Since the proper functions vanish exponentially at infinity, this integral is zero, and the orthogonality is demonstrated.

On account of the linearity of the wave equation any linear combination

$$\Psi = c_0 \Psi_0 + c_1 \Psi_1 + \dots c_k \Psi_k = c_0 \psi_0 e^{2\pi i E_0 / h t} + c_1 \psi_1 e^{2\pi i E_1 / h t} + \dots \quad (72)$$

is also a solution. Schrödinger interpreted this as an atomic state in which several natural frequencies are simultaneously developed, the

amplitude factors  $c_k$  being a measure of the excitation. If we use this solution to obtain the expression  $\Psi\bar{\Psi}$  we obtain

$$\begin{aligned}\Psi\bar{\Psi} = & c_0^2\psi_0^2 + c_1^2\psi_1^2 + \dots c_k^2\psi_k^2 + 2c_0c_1\psi_0\psi_1 \cos \frac{2\pi}{h}(E_1 - E_0)t + \dots \\ & + 2c_kc_l\psi_k\psi_l \cos \frac{2\pi}{h}(E_l - E_k)t. \quad . \quad . \quad . \quad (73)\end{aligned}$$

Thus the density of charge is composed of a part which is constant in time and another part whose magnitude oscillates with the frequencies

$$\nu_{kl} = \frac{E_l - E_k}{h}. \quad . \quad . \quad . \quad . \quad (74)$$

This variable charge, together with the nucleus, represents a variable dipole moment which, in accord with the laws of electrodynamics, emits light of frequencies  $\nu_{kl}$ . When averaged in time, the contributions of the variable terms to the charge density vanish, and since the total charge must be constant, the amplitude factors must satisfy the auxiliary condition

$$\sum c_k^2 = 1. \quad . \quad . \quad . \quad . \quad (75)$$

The dipole moment, from which the emission may be calculated according to p. 341, becomes

$$p_{kl} = -2c_kc_le \int r\psi_k\psi_ld\tau, \quad . \quad . \quad . \quad (76)$$

where  $r$  is the radius vector drawn from the nucleus to the volume element  $d\tau$ . The second Bohr postulate, by means of which the quantity  $h$  was introduced a second time without justification, is thus fulfilled automatically, on account of (74), in the Schrödinger formulation.

Although the Schrödinger interpretation must be replaced nowadays by a statistical view, it gives, nevertheless, a very vivid derivation of the dipole components which determine the intensity of spectral lines, viz.

$$P_{kl}^{(q)} = e \int q\psi_k\psi_ld\tau. \quad . \quad . \quad . \quad . \quad (76')$$

These components are identical with the "matrix elements" which determine the relative intensities of spectral lines in a quite different theory developed by Heisenberg.\* In particular, the matrix theory gives the old selection principles correctly.

The Schrödinger interpretation of the  $\Psi$  function met unsur-

\* W. Heisenberg, *Zeitschr. f. Phys.*, 33, p. 879 (1925). That the wave and matrix theories are mathematically equivalent was shown independently by Eckart and by Schrödinger.

mountable difficulties when an attempt was made to extend the reasoning to several electrons. In this case  $\psi$  depends upon  $3N$  co-ordinates, and the quantity  $\psi$  is determined, not as a function of position in actual space, but in the  $3N$ -dimensional configuration space. In place of the Schrödinger interpretation we then have a statistical one;  $\Psi\bar{\Psi}$  is a measure of the probability of finding an electron in the volume element  $d\tau$ , or—for more than one electron—the probability that the electrons lie in the  $3N$ -dimensional volume element of configuration space. We can then average over a large number of atoms and obtain a mean distribution of charge which corresponds to the Schrödinger assumption. The chief difference between the original formulation and this new one due to M. Born is, however, that, quite in keeping with the principles developed in § 2 (p. 690), we continue to think of the electrons as points and bring in the waves to take care of the indeterminacy. But since observation always gives an average over many atoms and fairly long times, we may in many cases treat the situation as though we were really dealing with the instantaneous distribution of charge for a single atom. The representations of charge density given by many recent papers are to be understood in this sense.

There is another difference in the two formulations that refers to the radiation process. According to (76), it is necessary that several proper functions be excited in order to attain a variable dipole moment. Translated into the language of statistics this means that in order to have emission, not all atoms can be in the same excited state. In reality the emission of radiation demands only that the excited state be occupied, and (76') then gives the transition probability, which is independent of the occupation of the lower energy states.

The continuous X-ray spectrum is also explained in this way by transitions between the closely spaced states for which  $E > 0$  (cf. p. 702). The largest jump yields a limit frequency  $\nu_L = E_0/h$ , where  $E_0$  is the initial energy of the electron.



## CHAPTER XLI

### ATOMIC PROBLEMS FIRST SOLVED BY THE WAVE MECHANICS

#### 1. The Theory of Perturbations in Wave Mechanics.

The wave mechanics not only yields the results of the Bohr theory—sometimes in an improved form, e.g. in the rotational term in band spectra—but succeeds in indicating the solutions of a number of important problems in face of which the Bohr theory was powerless. Of these we may mention the problem of several electrons (especially the helium problem), dispersion, non-polar chemical binding and, finally, the problem of nuclear disintegration (radioactivity). The concepts of wave mechanics throw new light on these questions.

A rigorous solution of the differential equation of the problem of  $n$  bodies is no more possible in wave mechanics than in celestial mechanics. In the former we must also apply methods of approximation, starting, just as in celestial mechanics, from the simple Kepler problem. The mathematical results needed for this perturbation theory will be developed in this section.

Let us add a perturbing term to the potential energy in the wave equation. For an atom, for example, this would be a series in descending powers of  $r$ , corresponding to a core potential. We therefore set

$$U = U^0 + U', \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

where  $U^0$  represents the potential energy of the unperturbed system. If  $U'$  is small, the proper values of the perturbed system will differ but little from those of the normal system, and we may therefore take

$$E_k = E_k^0 + \epsilon_k. \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

In the same way, we can relate the corresponding proper function to that of the normal system by setting

$$\psi = \psi_k^0 + \psi_k'. \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

We then have for the Schrödinger equation of the perturbed system

$$\Delta\psi_k^0 + \Delta\psi_k' + \frac{8\pi^2m_0}{h^2}(E_k^0 + \epsilon_k - U^0 - U')(\psi_k^0 + \psi_k') = 0. \quad (4)$$

Neglecting the products of perturbing terms and making use of the

fact that  $\psi_k^0$  satisfies the Schrödinger equation whose parameter is  $E_k$ , we obtain the following equation for  $\psi'$ :

$$\Delta\psi_k' + \frac{8\pi^2m_0}{h^2}(E_k^0 - U^0)\psi_k' = -\frac{8\pi^2m_0}{h^2}(\epsilon_k - U')\psi_k^0. \quad (5)$$

The left member of this non-homogeneous differential equation is identical with that of the equation for the normal system; hence for vanishing perturbation this leads both to the proper value  $E_k^0$  and also to the proper function  $\psi_k^0$ , as it must do. Now the theory of differential equations shows that a necessary condition for the existence of a solution of the non-homogeneous equation is the orthogonality of the perturbation function on the right to the solution of the homogeneous equation. This is proved by Green's Theorem, as at (67) and (68) (p. 704). Hence

$$\int(\epsilon_k - U')(\psi_k^0)^2 d\tau = 0. \quad . \quad . \quad . \quad . \quad (6)$$

But this is an equation determining the change in the proper value caused by the perturbation, which is what interests us primarily. Since  $\psi_k^0$  is normalized, this yields

$$\epsilon_k = \int U'(\psi_k^0)^2 d\tau. \quad . \quad . \quad . \quad . \quad (7)$$

This result corresponds to the theorem in celestial mechanics which states that, to a first approximation, the value of the perturbational energy is equal to the perturbation function averaged over the normal orbit.

If the perturbations of the proper values are known, the perturbing terms of the proper *functions* also may be determined by a development analogous to Fourier's series. That is, under the very general continuity conditions—which are always fulfilled in a physical problem—any function may be developed in a series of proper functions. If, for example, we set

$$f(x, y, z) = a_1\psi_1 + a_2\psi_2 + \dots a_k\psi_k + \dots, \quad . \quad (8)$$

we obtain (cf. p. 189) the coefficients exactly as for Fourier series by multiplying both members by  $\psi_m$  and integrating over the entire region. Because of the orthogonality, all terms with subscript different from  $m$  vanish, and there remains

$$\int f(x, y, z)\psi_m d\tau = a_m \int \psi_m^2 d\tau = a_m. \quad . \quad . \quad . \quad (9)$$

Accordingly, we develop the right side of equation (5) according to proper functions

$$-\frac{8\pi^2m_0}{h^2}(\epsilon_k - U')\psi_k^0 = \Sigma a_m\psi_m^0, \quad . \quad . \quad . \quad (10)$$

$$\text{where } a_m = \int \frac{8\pi^2 m_0}{h^2} (U' - \epsilon_k) \psi_k^0 \psi_m^0 d\tau \neq 0 \text{ for } k \neq m, \left. \begin{array}{l} \\ = 0 \text{ for } k = m. \end{array} \right\} \quad (11)$$

Similarly, we suppose the function  $\psi'$  to be written as a series of the type:

$$\psi'_k = \sum b_m \psi_m^0. \quad \cdot \cdot \cdot \cdot \cdot \cdot \quad (12)$$

These series are substituted in equation (5). Then, since the following relation holds for  $\psi_m^0$ :

$$\Delta \psi_m^0 - \frac{8\pi^2 m_0}{h^2} U^0 \psi_m^0 = -\frac{8\pi^2 m_0}{h^2} E_m^0,$$

comparison of the coefficients of each  $\psi_m$  on both sides of the equation yields

$$b_m = \frac{h^2}{8\pi^2 m_0} \frac{a_m}{(E_k^0 - E_m^0)}. \quad \cdot \cdot \cdot \cdot \cdot \quad (13)$$

Thus the condition of orthogonality, (6) and (11), prevents the occurrence of terms with vanishing denominator.

A complication arises here, as in celestial mechanics, in the very frequently occurring case in which the unperturbed system is degenerate, i.e. when several proper functions correspond to one and the same proper value. We say that the system has a degree of degeneracy  $(a - 1)$  if  $a$  linearly independent proper functions correspond to  $E_k^0$ . Every linear combination of these proper functions also represents a solution of the normal system. As may be seen from the derivation on p. 699, the proper functions corresponding to the same proper value are not, in general, orthogonal. However, by means of a linear substitution we can always set up linear combinations of the original proper functions in such a way that the proper functions corresponding to a single proper value also have the property of orthogonality; in fact, we still have a number of constants whose values we must determine when the perturbing force is applied. If we put

$$\left. \begin{array}{l} \overline{\psi}_k^1 = a_k^{11} \psi_k^1 + a_k^{12} \psi_k^2 + a_k^{13} \psi_k^3 + \dots + a_k^{1a} \psi_k^a \\ \overline{\psi}_k^2 = a_k^{21} \psi_k^1 + a_k^{22} \psi_k^2 + a_k^{23} \psi_k^3 + \dots + a_k^{2a} \psi_k^a \\ \overline{\psi}_k^a = a_k^{a1} \psi_k^1 + a_k^{a2} \psi_k^2 + a_k^{a3} \psi_k^3 + \dots + a_k^{aa} \psi_k^a \end{array} \right\}, \quad (14)$$

we have  $a^2$  coefficients, while the number of equations to be satisfied is  $a + a(a - 1)/2$ , viz.  $a$  equations involving  $\int (\overline{\psi}_k^i)^2 d\tau$  and  $a(a - 1)/2$  equations involving  $\int \overline{\psi}_k^i \overline{\psi}_k^j d\tau$ . Thus we have still  $a(a - 1)/2$  coefficients free. But this is the same as the number of independent



coefficients of an orthogonal linear substitution, i.e. a substitution whose coefficients  $b_k^{ih}$  satisfy the equations

$$\left. \begin{aligned} \sum b_k^{ih} b_k^{il} &= 0 \text{ for } h \neq l \\ &= 1 \text{ for } h = l \end{aligned} \right\} \quad \cdot \cdot \cdot \cdot \quad (15)$$

This transformation applied to the co-ordinates of a point signifies the rotation of a set of three rectangular axes. Actually, the orthogonal property of the proper functions is not disturbed by making such a substitution, as may be verified at once. We make the *a priori* assumption that the  $\alpha$  initial functions  $\psi_k^i$  satisfy the condition of orthogonality. If the degeneracy is removed by the perturbation, i.e. if the single proper value of multiplicity  $\alpha$  is split into  $\alpha$  different proper values  $E_0 + \epsilon_i$ , it is still undetermined which of the possible linear combinations of proper functions yield the proper functions of the perturbed system. We take this into account by forming the following linear combinations of the functions  $\overline{\psi_k^1}, \overline{\psi_k^2}, \overline{\psi_k^a}$ , which are already assumed to be orthogonal:

$$\left. \begin{aligned} \psi_k^{1*} &= b_k^{11} \overline{\psi_k^1} + b_k^{12} \overline{\psi_k^2} + \dots b_k^{1a} \overline{\psi_k^a} \\ \psi_k^{2*} &= b_k^{21} \overline{\psi_k^1} + b_k^{22} \overline{\psi_k^2} + \dots b_k^{2a} \overline{\psi_k^a} \\ &\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\ \psi_k^{a*} &= b_k^{a1} \overline{\psi_k^1} + b_k^{a2} \overline{\psi_k^2} + \dots b_k^{aa} \overline{\psi_k^a} \end{aligned} \right\} \quad \cdot \cdot \cdot \cdot \quad (16)$$

The coefficients are to satisfy the condition (15).

According to the theory of differential equations it is necessary, in order that a solution exist for the perturbed system, that the right side of equation (5) (p. 708) be orthogonal to *all* of the linearly independent solutions of the unperturbed system. If we insert a linear combination  $\psi_k^{i*}$  in the perturbation term, we obtain from the equations

$$\left. \begin{aligned} \int \overline{\psi_k^1} (\epsilon_k - U') \psi_k^{i*} d\tau &= 0 \\ \int \overline{\psi_k^2} (\epsilon_k - U') \psi_k^{i*} d\tau &= 0 \\ &\cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \\ \int \overline{\psi_k^a} (\epsilon_k - U') \psi_k^{i*} d\tau &= 0 \end{aligned} \right\} \quad \cdot \cdot \cdot \cdot \quad (17)$$

the  $\alpha$  proper values  $E + \epsilon_k$  as well as the  $\alpha$  initial functions (approximations of order zero). To show this we write the abbreviation

$$\int U' \overline{\psi_k^i} \overline{\psi_k^h} d\tau = \epsilon_k^{ih} = \epsilon_k^{hi}, \quad \cdot \cdot \cdot \cdot \quad (18)$$

and substitute the values of  $\psi_k^{i*}$  from (16) in (17). Then, since  $U'$  is assumed known as a function of position and since the functions

$\overline{\psi_k^h}$  are assumed known, we obtain a system of equations for the coefficients  $b_k^{ih}$ , which we may look upon as unknowns. The system

$$\left. \begin{aligned} b_k^{11}(\epsilon_k^{11} - \epsilon_k) + b_k^{12}\epsilon_k^{12} + \dots b_k^{1\alpha}\epsilon_k^{1\alpha} &= 0 \\ b_k^{11}\epsilon_k^{21} + b_k^{12}(\epsilon_k^{22} - \epsilon_k) + \dots b_k^{1\alpha}\epsilon_k^{2\alpha} &= 0 \\ . &. \\ b_k^{11}\epsilon_k^{\alpha 1} + b_k^{12}\epsilon_k^{\alpha 2} + \dots b_k^{1\alpha}(\epsilon_k^{\alpha\alpha} - \epsilon_k) &= 0 \end{aligned} \right\}. \quad (19)$$

corresponds to  $\psi_k^{10}$ . Since the equations are linear and homogeneous, the existence of a solution differing from zero requires that the determinant of the coefficients should vanish. This yields the so-called "secular equation"

$$\begin{vmatrix} (\epsilon_k^{11} - \epsilon_k) & \epsilon_k^{12} & \dots & \epsilon_k^{1a} \\ \epsilon_k^{21} & (\epsilon_k^{22} - \epsilon_k) & \dots & \epsilon_k^{2a} \\ \dots & \dots & \dots & \dots \\ \epsilon_k^{a1} & \epsilon_k^{a2} & \dots & (\epsilon_k^{aa} - \epsilon_k) \end{vmatrix} = 0. \quad (20)$$

This equation is of degree  $\alpha$  and so has  $\alpha$  roots. Because of the symmetry of the coefficients all roots are real. We again encounter this equation for each solution  $\psi_k^{(0)}$ . Substitution of one of the roots  $\epsilon_k^r$  yields the coefficients  $b_k^{(r)}$  after solving the system of equations. This then gives the zero order approximation to the perturbed proper value, from which the proper function corresponding to the value  $E_k^0 + \epsilon_k^r$  develops when the perturbation is applied to the system. If the secular equation has several equal roots, this means that the degeneracy has not been completely removed by the perturbation. Then, without the addition of further perturbations, each linear combination of the proper functions associated with the multiple roots again represents a solution. The new proper functions thus found form an orthogonal set because of the symmetry of the coefficient of the secular equation. This represents the multidimensional extension of the fact that the axes of the second-degree surfaces are mutually perpendicular (cf. *Ex. 18*, p. 37).

The perturbation theory in wave mechanics is confined to perturbing forces which have a potential. It is necessary to extend the Schrödinger equation in order to be able to compute the proper values for magnetic perturbations, i.e. those not derivable from a potential. Even in this instance we are able to derive only the normal Zeeman effect. Only by a very complicated generalization of the wave equation was it found possible to obtain the additional terms corresponding to electron spin in the theory based on models. This solution was given by Dirac, but cannot be discussed in detail here.



## 2. The Chemical Bond. Formation of the Hydrogen Molecule.

As an example of the application of quantum perturbation theory we deduce the solution of one of the most important problems in atomic physics—the nature of the chemical bond. There is little doubt regarding the nature of the forces holding together polar compounds, which are composed of positive and negative ions, as exemplified by salts. The forces in this instance are those of electrostatic attraction. A more difficult question is that concerning the cause of the limitation of the attractive forces, so that the oppositely charged ions remain at a definite distance from each other—in other words, what is the source of the repulsive forces which appear when the ions are brought close together? The wave mechanics offers information on this point. If we imagine the positive ions replaced by protons, they will, of course, be attracted by the negative ions, provided that the distance apart is not too small. Closer approach will ultimately cause the proton to enter the electron “cloud” surrounding the negative ion. The ions of a simple salt possess closed electron shells of the rare gas type, so that the electrons are symmetrically disposed. According to a fundamental law of the theory of the potential, only that part of the electron cloud which is within the surface of the sphere upon which the proton lies exerts any influence on the proton. Thus the proton will soon come to a point where the repulsion of the nuclear charge preponderates. Since, by p. 702, the density of charge decreases exponentially with distance from the centre, the repulsive forces must *increase* in this manner. Thus the simple law of force containing a single exponent is valid only at certain distances, and is merely a makeshift for the true law of force. This is manifested by the variations in the empirical value of the exponent of the repulsive force.

A much more difficult question to answer is that concerning the nature of the forces holding two atoms of the *same* kind together. The simplest instance of such a *non-polar* bond is given by  $H_2$ ; a more complicated instance is the bond between the carbon atoms in diamond. A method of explanation was first indicated by the wave mechanics through the introduction of *exchange forces* or, better, *exchange energies*. Heisenberg was led to this concept when computing the large difference in energy between the triplet and singlet terms of helium which—according to atom models—differ only in having the spins of the two electrons parallel or antiparallel. The small magnetic energy of two magnetons is not sufficient to account for the difference. If the electrons are distinguishable from each other, a given solution will always be accompanied by a second one belonging to the same characteristic value and in which the two electrons are interchanged. In this sense the system is degenerate and each linear combination also constitutes



a solution. Here again, as in § 1, we must seek those combinations from which the new characteristic values develop by the addition of a perturbation. The mutual electrostatic energy of the pair of electrons, which was previously neglected, constitutes the perturbation in this instance. Proceeding in this way, there results a large difference in the perturbations of the characteristic values for the two linear combinations that correspond to parallel and antiparallel spins. We can acquaint ourselves with the method by considering the bond between two hydrogen atoms:

Consider two hydrogen nuclei (protons),  $a$  and  $b$ , at a large distance  $R$  apart. Let there be also two electrons. Assume the distance of electron 1 from  $a$  to be  $r_{a1}$ —from  $b$ ,  $r_{b1}$ . Electron 2 is at a distance  $r_{a2}$  from  $a$ ,  $r_{b2}$  from  $b$ . Call the distance apart of the two electrons  $r_{12}$ . If the nuclei are held fixed, the Schrödinger equation for the system is

$$\frac{\partial^2 \psi}{\partial x_1^2} + \frac{\partial^2 \psi}{\partial y_1^2} + \frac{\partial^2 \psi}{\partial z_1^2} + \frac{\partial^2 \psi}{\partial x_2^2} + \frac{\partial^2 \psi}{\partial y_2^2} + \frac{\partial^2 \psi}{\partial z_2^2} + \frac{8\pi^2 m}{h^2} \left\{ E + \frac{e^2}{r_{a1}} + \frac{e^2}{r_{b2}} - \left( \frac{e^2}{R} + \frac{e^2}{r_{12}} - \frac{e^2}{r_{a2}} - \frac{e^2}{r_{b1}} \right) \right\} \psi = 0. \quad (21)$$

Assuming that electron 1 belongs to nucleus  $a$ , electron 2 to nucleus  $b$ , the quantities within parentheses are perturbing terms. Omitting these terms, we can arrange the equation as follows:

$$\Delta_1 \psi + \frac{8\pi^2 m}{h^2} \left( E_1 + \frac{e^2}{r_{a1}} \right) \psi + \Delta_2 \psi + \frac{8\pi^2 m}{h^2} \left( E_2 + \frac{e^2}{r_{b2}} \right) \psi = 0, \quad (22)$$

where

$$E = E_1 + E_2.$$

Each pair of terms set equal to zero represents the Schrödinger equation of the hydrogen atom. If, according to equation (60) on p. 702, we therefore take the normalized solutions for the hydrogen atom in the ground states ( $n = 1, l = 0, m = 0$ ) for the nucleus  $a$  as well as for the nucleus  $b$  to be

$$\left. \begin{aligned} \psi_a(1) &= \frac{1}{\sqrt{\pi}} \left( \frac{1}{r_0} \right)^{3/2} e^{-r_{a1}/r_0} \\ \psi_b(2) &= \frac{1}{\sqrt{\pi}} \left( \frac{1}{r_0} \right)^{3/2} e^{-r_{b2}/r_0} \end{aligned} \right\}, \quad \cdot \cdot \cdot \cdot \quad (23)$$

then

$$\psi = \psi_a(1) \psi_b(2), \quad \cdot \cdot \cdot \cdot \cdot \cdot \quad (24)$$

with  $E_1 = E_2 = E_0$ , i.e.  $E = 2E_0$ , certainly is a solution of the equation of the combined system from which the interactions have been

omitted.\* This may be verified at once by substitution in equation (22). But we can assume with equal right that electron 1 belongs to nucleus  $b$  and electron 2 belongs to nucleus  $a$ . This leads to a second solution which corresponds to the same characteristic value  $2E_0$ , and which has the form

$$\psi = \psi_a(2)\psi_b(1). \quad . \quad . \quad . \quad . \quad . \quad (25)$$

The original problem—that of two hydrogen atoms assumed to constitute a single system—is thus singly degenerate; each linear combination  $\psi_0'$  and  $\psi_0''$  of the two proper functions also represents a solution:

$$\left. \begin{aligned} \psi_0' &= \alpha\psi_a(1)\psi_b(2) + \beta\psi_a(2)\psi_b(1) \\ \psi_0'' &= \gamma\psi_a(1)\psi_b(2) + \delta\psi_a(2)\psi_b(1) \end{aligned} \right\} . \quad . \quad . \quad (26)$$

Four coefficients are at our disposal. Three of these are determined by the condition of orthogonality and the normalization to unity:

$$\int \psi_0'^2 d\tau = \int \psi_0''^2 d\tau = 1; \quad \int \psi_0'\psi_0'' d\tau = 0. \quad . \quad (27)$$

According to the procedure in § 1 (p. 707), the fourth constant and the correction term  $\epsilon$  in the energy are determined by the condition that, after inserting the first approximation  $\psi_0 + \psi_1$ , the right member of the resulting non-homogeneous equation must be orthogonal to all solutions of the homogeneous equation. Since we have two solutions here, the secular equation is quadratic in  $\epsilon$ , yielding two roots which we designate by  $\epsilon_s$  and  $\epsilon_a$ . The approximation of zero order corresponding to  $\epsilon_s$ , from which the perturbed proper functions are developed, results after a lengthy calculation:

$$\psi_0^s = \frac{1}{\sqrt{2 + 2S}} \{ \psi_a(1)\psi_b(2) + \psi_a(2)\psi_b(1) \}; \quad . \quad (28)$$

\* The value of the multiplicative constant  $K$  corresponding to normalization to unity is not to be obtained from equation (60). It must be separately computed from the relation

$$\int K^2 \psi^2 d\tau = K^2 \int_0^{4\pi} \int_0^\infty e^{-2r/r_0} r^2 d\Omega dr = 1,$$

from which it follows that

$$K = \frac{1}{\sqrt{\pi}} \left( \frac{1}{r_0} \right)^{3/2}.$$

that corresponding to  $\epsilon_a$  is

$$\psi_0^a = \frac{1}{\sqrt{2 + 2S}} \{ \psi_a(1)\psi_b(2) - \psi_a(2)\psi_b(1) \}, \quad (29)$$

where  $S$  is an abbreviation for the volume integral

$$\int \psi_a(1)\psi_b(2)\psi_a(2)\psi_b(1) d\tau.$$

Thus the proper function which is associated with  $\epsilon_s$  is symmetric in both electrons, i.e. it is not altered by interchanging 1 and 2; the

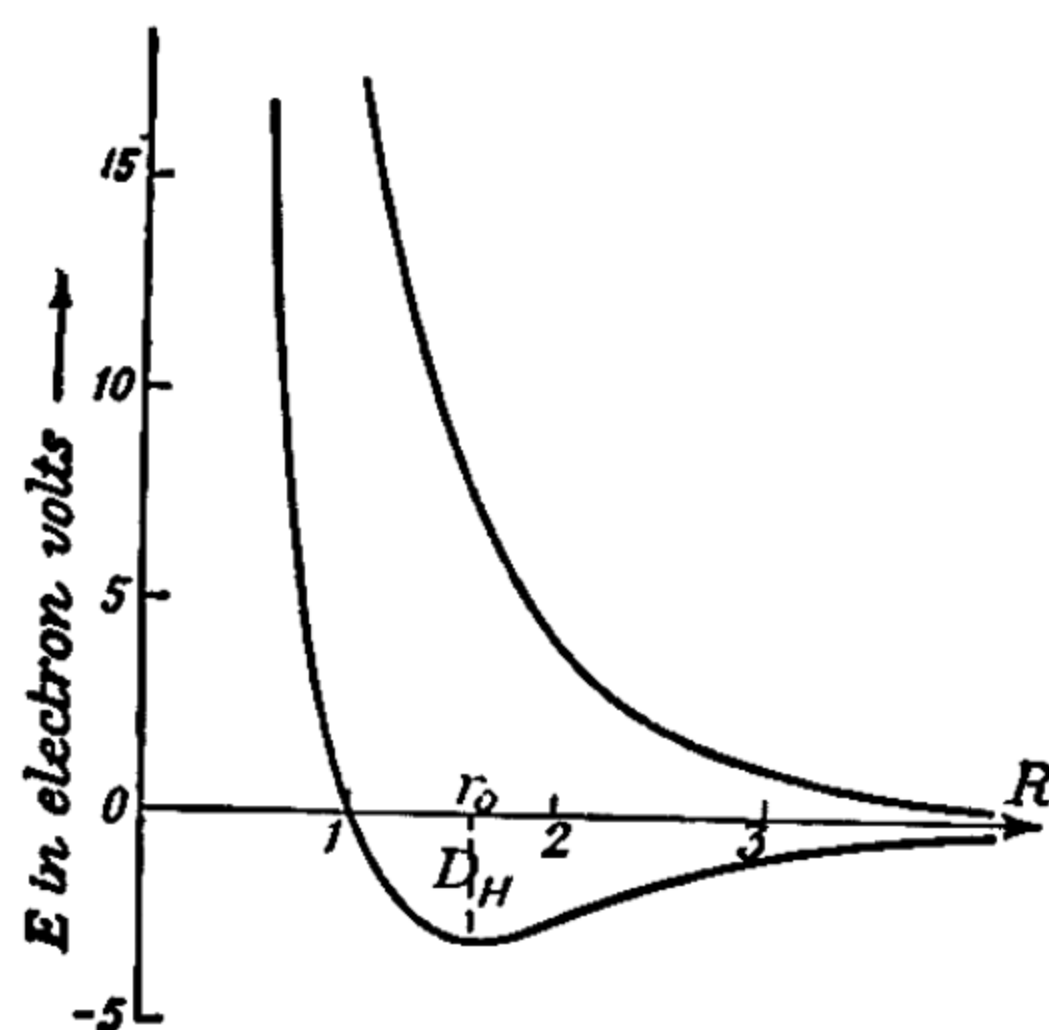


Fig. 1

other solution—that corresponding to  $\epsilon_a$ —is antisymmetric, i.e. its sign changes if the two electrons are interchanged. The computation of the roots of the secular equation is somewhat troublesome, so that only the results are given here in the form of two curves which exhibit the value of  $\epsilon$  as a function of the parameter  $R$  (fig. 1). The important feature is that the symmetric solution has an energy minimum at a definite distance from the nucleus. This solution therefore corresponds to

the *chemical bond*, while the other solution, for which the additional energy is always positive, corresponds to elastic rebound.

The wave mechanics was first applied in this manner by Heitler and London.\* It gives a surprisingly accurate value of the heat of reaction in this case, considering that only the zero order approximation was computed. But in addition, this method is capable of yielding information on the possibility of forming certain compounds if we translate the Pauli Principle into the language of wave mechanics. According to this law, no two electrons in an atom can agree in all four quantum numbers. In wave mechanics, this means that the  $\psi$  function cannot be symmetric in the co-ordinates of two electrons. This might cause one to believe that the symmetric solution for the hydrogen molecule is excluded. It must be remembered, however, that the Pauli Principle includes the spin orientation (parallel or antiparallel to a given direction) as a fourth co-ordinate, while up to this point the electron spin has not been included in the wave me-

\* G. Heitler and F. London, *Zeitschr. f. Phys.*, 44, p. 455 (1927).



chanics. According to Pauli's rule, two electrons having opposite spins may agree in the remaining quantum numbers. Thus, taking into account the doubling of the possibilities resulting from the spin, we can adopt Pauli's Principle by saying that the  $\psi$  function of an atom or molecule can at most be symmetric in certain *pairs* of electrons, i.e. it may be such that when electron  $x$  and electron  $y$  are interchanged the function remains unaltered, but in that case not when  $x$  and electron  $z$  are interchanged. The electrons in which the  $\psi$  function is symmetric have opposite directions of spin. For the helium atom, for example, the proper function for the ground state is symmetric in both electrons, corresponding to the equal quantum numbers  $n$ ,  $l$  and  $m_l$ . The symmetric function for two atoms would then be symmetric in four electrons, which is impossible according to Pauli's Principle. But the solution corresponds to that of the chemical bond even in this case. *The symmetric proper function resulting from the combination of two electrons which, in the individual atoms, have no partners whose three quantum numbers are the same, but which have oppositely oriented spins, represents the interpretation of chemical valence on the wave mechanics.*

### 3. The Theory of Dispersion and of the Raman Effect in Wave Mechanics.

We now consider the behaviour of an electron in the rapidly alternating electric field of a light wave. Here, as in the classical dispersion theory, we neglect the magnetic forces. The perturbing potential contains the time explicitly. The previous form of the wave equation is not suitable for treating perturbation problems of this kind, for it already contains the characteristic parameter, and hence there is a separate differential equation for each state. It is not difficult, however, to eliminate the parameter  $E$ . It follows from our solution

$$\Psi = \psi e^{2\pi i E/h t}, \quad . . . . . (30)$$

that

$$\frac{\partial \Psi}{\partial t} = \frac{2\pi i E}{h} \Psi; \quad . . . . . (31)$$

hence

$$E = \frac{h}{2\pi i} \frac{\partial \Psi}{\partial t} \cdot \frac{1}{\Psi} \quad . . . . . (32)$$

Substituting this in the actual wave equation and remembering that

$$\frac{\partial^2 \Psi}{\partial t^2} = - \frac{4\pi^2 E^2}{h^2} \Psi, \quad . . . . . (33)$$

we obtain from

$$\Delta \Psi = \frac{2m(E - U)}{E^2} \frac{\partial^2 \Psi}{\partial t^2}$$

the form 
$$\Delta\Psi - \frac{8\pi^2m}{h^2} U\Psi - \frac{4\pi im}{h} \frac{\partial\Psi}{\partial t} = 0, \quad . \quad . \quad . \quad (34)$$

which no longer contains a characteristic parameter.

Let the electric field of the incident light wave be along the  $z$ -axis and of amplitude  $A$ . The perturbing potential is then

$$U' = -eAz \cos 2\pi\nu t, \quad . \quad . \quad . \quad (35)$$

and the new form of the wave equation is

$$\Delta\Psi - \frac{8\pi^2m}{h^2} (U^0 - eAz \cos 2\pi\nu t)\Psi - \frac{4\pi im}{h} \frac{\partial\Psi}{\partial t} = 0,$$

or

$$\Delta\Psi - \frac{8\pi^2m}{h^2} U^0\Psi - \frac{4\pi im}{h} \frac{\partial\Psi}{\partial t} = -\frac{4\pi^2m}{h^2} eAz (e^{2\pi i\nu t} + e^{-2\pi i\nu t})\Psi. \quad (36)$$

We may substitute the unperturbed  $k$ th proper function  $\Psi_k^0 = \psi_k^0 e^{2\pi i\nu_k t}$  for  $\Psi$  in the perturbing term on the right.\* We then have

$$\begin{aligned} \Delta\Psi - \frac{8\pi^2m}{h^2} U^0\Psi - \frac{4\pi im}{h} \frac{\partial\Psi}{\partial t} \\ = -\frac{4\pi^2m}{h^2} eAz (e^{2\pi i(\nu_k + \nu)t} + e^{2\pi i(\nu_k - \nu)t}) \psi_k^0. \quad . \quad (37) \end{aligned}$$

It is natural to take as a solution the sum of the normal solution and two other terms of frequency  $\nu_k + \nu$  and  $\nu_k - \nu$  respectively, i.e.

$$\Psi = \Psi_k^0 + \psi_+ e^{2\pi i(\nu_k + \nu)t} + \psi_- e^{2\pi i(\nu_k - \nu)t}. \quad . \quad (38)$$

Substituting this in equation (37) and noticing that  $\Psi_k^0$  causes the left member to vanish, we obtain two differential equations—one for  $\psi_+$ , one for  $\psi_-$ —which we may write as the single equation

$$\Delta\psi_{\pm} + \frac{8\pi^2m}{h^2} (E_k \pm h\nu - U^0)\psi_{\pm} = -\frac{4\pi^2m}{h^2} eAz \psi_k^0. \quad (39)$$

This equation is again of the form of the perturbed Schrödinger equation which was treated in § 1 (p. 707). It differs from the former equation, however, by the presence of the term  $\pm h\nu$ . This has an important consequence: in the case at hand, the homogeneous equation has no solution whatsoever. The parameter is  $E_k \pm h\nu$ , and a solution exists only for the value  $E_k$ . The theory of differential equations shows that the *non-homogeneous* equation always has a solution in this case.

\* Contrary to our procedure in § 1, we now omit the superscript 0 on  $E_k^0$  and  $\nu_k^0$ , since there can be no doubt here that these quantities refer to the unperturbed system.

The condition of orthogonality which must be satisfied when the homogeneous equation has a solution, in order to prevent the appearance of the term with vanishing denominator  $E_k - E_k$ , is absent here. As for the rest, we can then proceed according to the method developed in § 1: we expand the function  $z\psi_k^0$  of the right member in a series involving the proper functions

$$z\psi_k^0 = \sum a_{nk}\psi_n^0, \quad . \quad . \quad . \quad . \quad . \quad (40)$$

$$a_{nk} = + \int z\psi_k^0\psi_n^0 d\tau, \quad . \quad . \quad . \quad . \quad . \quad (41)$$

and write  $\psi_+$  and  $\psi_-$  also as series of the same type:

$$\psi_{\pm} = \sum b_{\pm n}\psi_n^0. \quad . \quad . \quad . \quad . \quad . \quad (42)$$

This is substituted in (39). Remembering that  $\psi_n^0$  satisfies the equation

$$\Delta\psi_n^0 + \frac{8\pi^2m}{h^2}(E_n - U^0)\psi_n^0 = 0, \quad . \quad . \quad . \quad (43)$$

comparison of coefficients yields

$$b_{\pm n} = -\frac{1}{2} \frac{eAa_{nk}}{E_k - E_n \pm h\nu}. \quad . \quad . \quad . \quad . \quad (44)$$

Thus we can write as the complete solution

$$\Psi_k = \psi_k^0 e^{2\pi i\nu_k t} - \frac{1}{2} \sum_n \psi_n^0 eAa_{nk} \left\{ \frac{e^{2\pi i(\nu_k + \nu)t}}{E_k - E_n + h\nu} + \frac{e^{2\pi i(\nu_k - \nu)t}}{E_k - E_n - h\nu} \right\}. \quad (45)$$

By using our working hypothesis that  $e\Psi\bar{\Psi}$  gives the density of charge, we can obtain the electric moment of the atom from this equation. The polarization is concerned with the  $z$ -component:

$$P_z = e \int z\Psi\bar{\Psi} d\tau. \quad . \quad . \quad . \quad . \quad . \quad (46)$$

Substituting the value resulting from equation (45), and using the relation

$$\int z\psi_n\psi_k d\tau = a_{nk},$$

we obtain the following expression for the temporally variable part of the moment (the constant part is of no further interest):

$$P_z(t) = 2e^2A \cos 2\pi\nu t \sum_n a_{nk}^2 \frac{E_n - E_k}{(E_n - E_k)^2 - h^2\nu^2}. \quad . \quad (47)$$

Thus the temporally variable part has the same frequency as the incident wave. Since the ratio of the amplitude of the electric moment



to that of the electric field strength is the polarizability  $\alpha$ , this quantity is thus determined.

If this value is substituted in equation (3') (p. 452), the value obtained for the index of refraction in the state  $k$  is

$$\begin{aligned} n - 1 &= 4\pi N e^2 \sum_n a_{nk}^2 \frac{E_n - E_k}{(E_n - E_k)^2 - h^2 \nu^2} \\ &= 4\pi N \frac{e^2}{h} \sum_n a_{nk}^2 \frac{\nu_{nk}}{\nu_{nk}^2 - \nu^2}. \quad \cdot \cdot \cdot \cdot (48) \end{aligned}$$

If we notice, in addition, that the spectral *lines* of the atom are given by the frequencies  $\nu_{kn}$ , we see that we have obtained the characteristic denominator  $\nu_{nk}^2 - \nu^2$  of the classical theory of dispersion by means of quantum mechanics. Thus, apart from the fact that the constants in the numerator have a different meaning, we have here the connecting link with the old dispersion formula of p. 452.

In the language of the Bohr theory, the quantity  $a_{nk}$  is a measure of the probability of a transition connected with radiation of frequency  $\nu_{nk}$  for a spatial distribution of radiant energy as given by an oscillator vibrating parallel to the  $z$ -axis. Thus  $a_{nk}$  is a measure of the strength of the classical equivalent oscillator. Comparing the constants of the classical and wave mechanics dispersion formulæ we obtain

$$f_{nk} = \frac{8\pi^2 m a_{nk}^2 \nu_{nk}}{h}. \quad \cdot \cdot \cdot \cdot (48')$$

The terms having the negative sign were not included in the classical formula. These appear if  $E_n < E_k$ . For this it is necessary for the substance examined to be in the excited state, while dispersion measurements are usually made on substances in the ground state. Using excited hydrogen atoms, Ladenburg succeeded in establishing the existence of the negative dispersion terms.

The remaining components of the induced electric moment are of no consequence for the dispersion, but determine—together with the  $z$ -component—the *scattering power* of the atom with respect to the incident radiation. We shall not consider the general theory of the scattering of light by individual atoms, but shall treat the particularly important special case of the *Raman effect*.

If several proper functions of a system are active, i.e. if the system is in a state of transition, combination frequencies will be observed in the scattered radiation. This effect was predicted theoretically by Smekal, and found independently by experiment by Raman and by Landsberg and Mandelstam at about the same time. Let the normal system contain the frequencies  $\nu_k$  and  $\nu_l$ . Then, by means of the same perturbation theory as above, the solution for the

perturbed system when a wave of frequency  $\nu$  is incident on the system is \*

$$\begin{aligned}\Psi = & \psi_k^0 e^{2\pi i \nu_k t} + \psi_l^0 e^{2\pi i \nu_l t} \\ & - \frac{1}{2} \sum \psi_n^0 e^2 A a_{nk} \left\{ \frac{e^{2\pi i (\nu_k + \nu) t}}{E_k - E_n + h\nu} + \frac{e^{2\pi i (\nu_k - \nu) t}}{E_k - E_n - h\nu} \right\} \\ & - \frac{1}{2} \sum \psi_n^0 e^2 A a_{nl} \left\{ \frac{e^{2\pi i (\nu_l + \nu) t}}{E_l - E_n + h\nu} + \frac{e^{2\pi i (\nu_l - \nu) t}}{E_l - E_n - h\nu} \right\}. \quad (49)\end{aligned}$$

The conjugate solution is

$$\begin{aligned}\bar{\Psi} = & \psi_k^0 e^{-2\pi i \nu_k t} + \psi_l^0 e^{-2\pi i \nu_l t} \\ & + \frac{1}{2} \sum \psi_n^0 e^2 A a_{nk} \left\{ \frac{e^{-2\pi i (\nu_k + \nu) t}}{E_k - E_n + h\nu} + \frac{e^{-2\pi i (\nu_k - \nu) t}}{E_k - E_n - h\nu} \right\} \\ & + \frac{1}{2} \sum \psi_n^0 e^2 A a_{nl} \left\{ \frac{e^{-2\pi i (\nu_l + \nu) t}}{E_l - E_n + h\nu} + \frac{e^{-2\pi i (\nu_l - \nu) t}}{E_l - E_n - h\nu} \right\}. \quad (49')\end{aligned}$$

The scattering is determined by the moments

$$M_x = \int x \Psi \bar{\Psi} d\tau, \quad M_y = \int y \Psi \bar{\Psi} d\tau, \quad M_z = \int z \Psi \bar{\Psi} d\tau. \quad (50)$$

It is immediately evident that, in forming these expressions, terms of frequency  $\nu - \nu_{lk}$  and  $\nu + \nu_{lk}$  are obtained, i.e. the effect of the scattering is the same as if a quantum of the characteristic frequency  $\nu_{lk}$  were added to or subtracted from the quantum  $h\nu$  of the incident light. The constants appearing when the moment is determined are of great importance. If, for example, we compute  $M_y$ , and use, in addition to

$$a_{nk} = \int z \psi_n \psi_k d\tau, \quad . \quad . \quad . \quad . \quad (51)$$

the abbreviation

$$b_{nk} = \int y \psi_n \psi_k d\tau, \quad . \quad . \quad . \quad . \quad (52)$$

then the term containing  $\nu \pm \nu_{lk}$  has coefficients which contain the products  $b_{nk} a_{nl}$ . This implies that the intensity of a Raman scattered ray is *not* determined by the intensity of the light of frequency  $\nu_{lk}$  but by the intensities of lines resulting from transitions from higher levels  $n$  to the levels  $k$  and  $l$ . Thus is explained the fact that lines which are absent from the infra-red spectrum of a molecule on account of the vanishing of  $a_{lk}$  can nevertheless give rise to intensive Raman lines.

\* On account of the smallness of the perturbation we can neglect all products of perturbing terms.

The theory developed here is still incomplete to the extent that it predicts equal intensity for the two lines  $\nu + \nu_{ik}$  and  $\nu - \nu_{ik}$ , and connects the appearance of Raman lines with excitation of the molecule. Actually, the "Stokes" line  $\nu - \nu_{ik}$  can appear even at absolute zero, while the "anti-Stokes" line  $\nu + \nu_{ik}$  only sets in upon excitation, i.e. at higher temperatures.\* This is readily understood from the standpoint of quanta, for at absolute zero a quantum of radiation can transmit energy to the scattering molecule but can obtain no energy from it.

The questions of intensity and of polarization have been solved by a more refined theory of the effect. It is possible to gain an insight into the nature of what is going on by means of the following considerations which are in the spirit of the actual calculation: In questions dealing with dispersion, the classical concepts generally prove to be far-reaching in their validity. Classically, the scattering of light arises from the forced vibration of the electron shells, with consequent radiation. If the polarizability of, say, a diatomic molecule changes with the nuclear separation, then the amplitude of the radiation will vary with the same frequency as the nuclear vibration; but a spectroscope analyses this modulated wave into its component frequencies, as detailed on p. 58.

#### 4. The Surmounting of Potential Barriers in the Wave Mechanics.

The phenomenon of radioactive disintegration, which will be considered in Chapter XLII (p. 739), presents us with the following problem.

By experiments on  $\alpha$ -particle scattering it has been shown that for Uranium 1 the Coulomb field holds down to a distance of  $3 \times 10^{-12}$  cm. Apart from a nuclear charge which is two units less, the same potential curve is valid also for the element UX<sub>1</sub> which results by the ejection of an  $\alpha$ -particle. For a closer approach, the potential curve must finally bend downward, for in the U1 atom the  $\alpha$ -particle is generally bound, and hence must be at a minimum of the potential curve (fig. 2). Again, an  $\alpha$ -particle ejected by U1 has an amount of energy which corresponds to its being released from rest at a distance  $r_1 = 6 \times 10^{-12}$  cm. and being allowed to accelerate in the field of the nucleus. The question arises: How does the  $\alpha$ -particle cross the potential barrier between its normal position and the point  $r_1$ ? There is no possibility of explaining this fact on the basis of classical mechanics. Gamow and, independently, Condon first recognized that in wave mechanics it is possible to encounter a particle where it does not belong on the basis of its classical energy—i.e. every potential barrier is surmountable. Indeed, according to the statistical inter-

\* This terminology is based on Stokes' law of fluorescence, according to which fluorescent radiation is always of longer wave-length than the exciting light. The law follows at once from a consideration of absorption and emission in terms of an energy-level diagram. If the atom or molecule is not in an excited state to begin with, the energy of the re-radiated photon can never be greater than that of the one that was absorbed.



pretation of the Schrödinger  $\psi$  function there is a definite probability that the electron of the hydrogen atom will be found far from the nucleus at a place where it could not be expected at all on the basis of its energy (proper value) parameter.

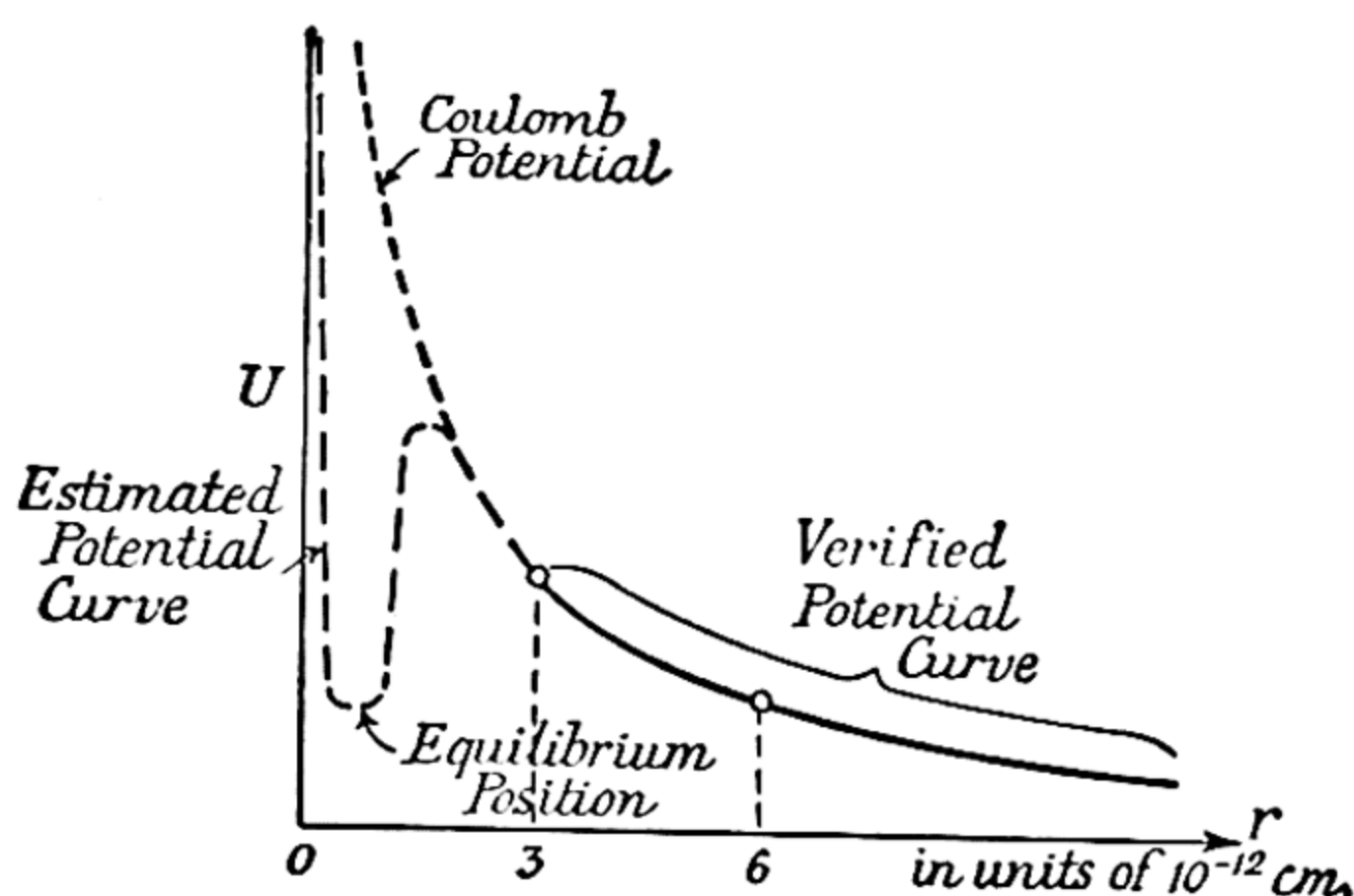


Fig. 2

We shall consider the surmounting of the potential barrier for the simple model investigated by Gamow, since we are interested only in the essential features here. Assume the potential energy of a particle to have a constant value  $U$  from  $x = 0$  to  $x = l$  and to vanish in the remainder of the region. Now allow a pencil of  $\alpha$ -particles (or, equally well, of electrons) to pass from left to right. According to classical mechanics, all particles whose kinetic energy is smaller than  $U$  will be thrown back, while all particles possessing greater energy will pass the barrier. According to the quantum mechanics, neither of these statements is exactly true. We limit our present considerations to the determination of the number of particles with smaller energy which succeed in crossing the obstacle. To this end we must divide the space into three regions  $x < 0$ ,  $0 < x < l$ ,  $x > l$ , and assume a Schrödinger wave for each. The boundary conditions must also be taken into account.

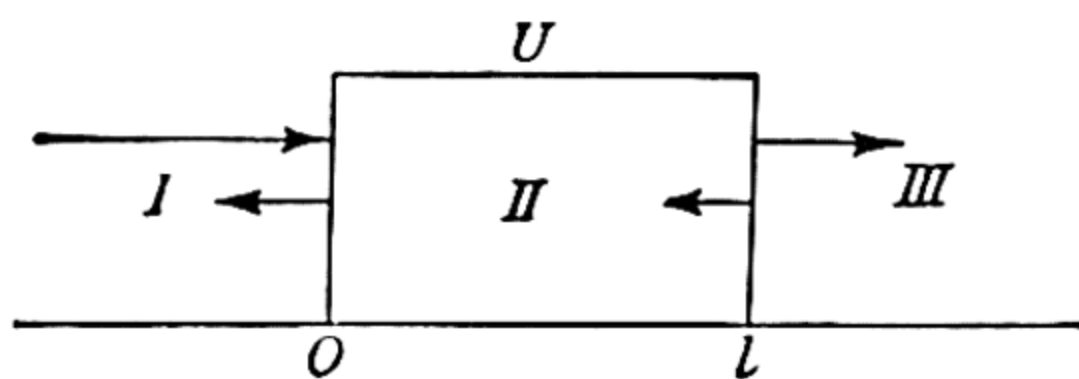


Fig. 3

The situation is simplest in the third region; there we certainly have but one wave travelling to the right, while partial reflections are to be expected at the boundaries, so that we must have waves moving in both directions in the other two regions (cf. fig. 3). The Schrödinger equation in Regions I and III is

$$\frac{d^2\psi}{dx^2} + \frac{8\pi^2M}{h^2} E\psi = 0, \quad \dots \dots \dots (53)$$

while in Region II

$$\frac{d^2\psi}{dx^2} + \frac{8\pi^2M}{h^2} (E - U)\psi = 0. \quad \dots \quad (53')$$

Since we are concerned only with the ratios of the amplitudes, we set the amplitude in Region III equal to unity, and thus obtain the three solutions for the three regions:

$$\left. \begin{aligned} \psi_1 &= a_1 e^{ik_0 x} + b_1 e^{-ik_0 x} \\ \psi_2 &= a_2 e^{ik_1 x} + b_2 e^{-ik_1 x} \\ \psi_3 &= e^{-ik_1 x} \end{aligned} \right\}, \quad \dots \quad (54)$$

where

$$k_0 = \frac{2\pi}{h} \sqrt{2ME}, \quad k_2 = \frac{2\pi}{h} \sqrt{2M(E - U)}.$$

The solutions must merge continuously at the boundaries in such way that at each boundary the number of arriving particles must, on the average, be equal to the sum of the number reflected and the number transmitted. This signifies the continuity of the  $\psi$  function. On account of the existence of the Schrödinger equation the first derivative must be continuous also, otherwise  $\Delta\psi$  and hence  $\psi$  would be infinite at the boundaries. These conditions yield the following equations at  $x = l$ :

$$\left. \begin{aligned} e^{-ik_1 l} &= a_2 e^{ik_1 l} + b_2 e^{-ik_1 l} \\ \frac{k_0}{k_2} e^{-ik_1 l} &= -a_2 e^{ik_1 l} + b_2 e^{-ik_1 l} \end{aligned} \right\}. \quad \dots \quad (55)$$

Solved for  $a_2$  and  $b_2$  these become

$$\begin{aligned} a_2 &= \frac{1}{2} \left( 1 - \frac{k_0}{k_2} \right) e^{-i(k_0 + k_1)l}, \\ b_2 &= \frac{1}{2} \left( 1 + \frac{k_0}{k_2} \right) e^{-i(k_0 - k_1)l}. \end{aligned}$$

Similarly, at  $x = 0$  we obtain

$$\left. \begin{aligned} a_1 &= \frac{1}{2} \left( 1 + \frac{k_2}{k_0} \right) a_2 + \frac{1}{2} \left( 1 - \frac{k_2}{k_0} \right) b_2 \\ b_1 &= \frac{1}{2} \left( 1 - \frac{k_2}{k_0} \right) a_2 + \frac{1}{2} \left( 1 + \frac{k_2}{k_0} \right) b_2 \end{aligned} \right\}. \quad \dots \quad (56)$$

We are primarily interested in  $b_1$ , i.e. the amplitude of the incident wave relative to that of the wave traversing Region III. Substituting the values of  $a_2$  and  $b_2$  from (55) in (56), there results

$$b_1 = \frac{1}{4} \left(1 - \frac{k_2}{k_0}\right) \left(1 - \frac{k_0}{k_2}\right) e^{-i(k_0 + k_2)l} \\ + \frac{1}{4} \left(1 + \frac{k_2}{k_0}\right) \left(1 + \frac{k_0}{k_2}\right) e^{-i(k_0 - k_2)l}. \quad \cdot \cdot \cdot (57)$$

The *intensity* of the wave is obtained by multiplying the amplitude by its complex conjugate value. For  $E > U$  this yields

$$b_1 \bar{b}_1 = \frac{1}{16} \left(1 - \frac{k_2}{k_0}\right)^2 \left(1 - \frac{k_0}{k_2}\right)^2 + \frac{1}{16} \left(1 + \frac{k_2}{k_0}\right)^2 \left(1 + \frac{k_0}{k_2}\right)^2 \\ + \frac{1}{8} \left[1 - \left(\frac{k_0}{k_2}\right)^2\right] \left[1 - \left(\frac{k_2}{k_0}\right)^2\right] \cos 2k_2 l. \quad \cdot \cdot \cdot \cdot (58)$$

According to classical mechanics, *all* particles would reach Region III when  $E > U$ ; the above formula gives the ratio of the number of incident particles to the number of particles passing through as given by wave mechanics. The case  $E < U$  is, however, of greater importance. In this case we take

$$k_2 = ik = \frac{2\pi i}{h} \sqrt{2M(U - E)} \quad \cdot \cdot \cdot (59)$$

and obtain

$$b_1 \bar{b}_1 = \frac{1}{2} - \frac{1}{8} \left(\frac{k_0}{k} - \frac{k}{k_0}\right)^2 + \frac{1}{8} \left(\frac{k_0}{k} + \frac{k}{k_0}\right)^2 \cosh 2kl. \quad (60)$$

Since  $\cosh x$  behaves like  $e^x$  when  $x$  is large, we arrive at the essential feature of the connexion between energy and radioactive constant in spite of the simplicity of the model used. The intensity of the incident wave, referred to that of the wave in Region III, increases exponentially with the difference  $U - E$ , or—reversing the statement—the ratio of the number of particles which succeed in crossing the barrier to the number arriving there decreases exponentially with increasing  $U - E$ .

The result has an importance reaching far beyond the special problem from which it has been derived here. The overcoming of obstacles in quantum mechanics which appear insurmountable in ordinary mechanics is important in considering such problems as the escape of electrons from metallic surfaces and the kinetics of chemical reactions, but these questions cannot be entered into in this work.



## 5. Energy Bands of Electrons in Metals.

### (a) General Survey of the Problem

The concept of an electron gas moving freely among the ions of the lattice, when used in conjunction with the Fermi statistics, is able to account correctly for many features of the properties of metals (see p. 602). As mentioned there, however, further development of the theory must take account of the interaction of the electrons with the lattice; we can no more neglect the large electrostatic forces here than we could in dealing with electrolytes. In this way only we arrive at an understanding of phenomena which were formerly not even qualitatively explained, e.g. the reversed sign of the Hall effect occurring for some metals (cf. p. 733). Further, such a theory would be expected to furnish a criterion as to whether or not a given lattice made up of atoms of a specified kind will possess metallic properties. These ques-

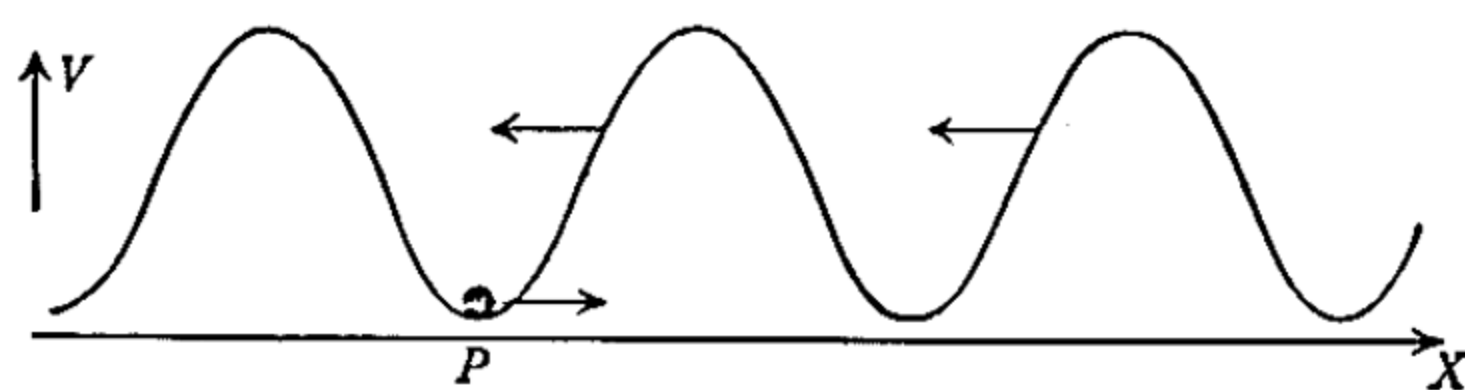


Fig. 4

tions are actually answered by incorporating the Fermi statistics into a wave-mechanical treatment of the motion of electrons in the periodic potential field of the lattice. Because of the singular behaviour of particles with respect to potential barriers (cf. § 4), this motion differs from that to be expected on the basis of particle mechanics. This is most easily seen in the one-dimensional case, to which we restrict our treatment. This calculation involves less work than the one in three dimensions, but qualitatively—and that is all that concerns us for the present—it yields the same result.

Suppose an electron moving in the direction of the positive  $x$ -axis (fig. 4) to be momentarily in the potential trough at P. According to quantum mechanics it can surmount the next crest even if its kinetic energy is less than the amount corresponding to the crest. On the other hand, even in instances where it has sufficient energy, there exists a given—although slight—possibility of reflection, as mentioned above. If the de Broglie waves reflected from the several crests agree in phase, the result may be a reflection of such magnitude that the motion to the right becomes impossible. This is nothing more nor less than the reflection in depth for X-rays, which was treated in detail on p. 388. Precisely this circumstance is the key to an understanding of the original questions.

(b) *Wave-Mechanical Expression for Current Density*

On p. 703 the expression  $\Psi\bar{\Psi}$  was shown to be a measure of the density of *charge*. Now an analogous expression for the *current* density will be deduced. Since a current is a process occurring in time, we must start from the Schrödinger equation (p. 717), which depends on the time:

$$\Delta\Psi - \frac{8\pi^2m}{h^2} U\Psi - \frac{4\pi im}{h} \frac{\partial\Psi}{\partial t} = 0. \quad . \quad . \quad . \quad (63)$$

According to a theorem in the theory of functions of a complex variable, the conjugate function  $\bar{\Psi}$  satisfies an equation in which all coefficients are replaced by their conjugates, since the real and the imaginary parts must separately satisfy the same equation. Hence  $\bar{\Psi}$  satisfies the equation

$$\Delta\bar{\Psi} - \frac{8\pi^2m}{h^2} U\bar{\Psi} + \frac{4\pi im}{h} \frac{\partial\bar{\Psi}}{\partial t} = 0. \quad . \quad . \quad . \quad (64)$$

Multiplying (63) by  $\bar{\Psi}$  and (64) by  $\Psi$  and subtracting, we have

$$\frac{4\pi im}{h} \frac{\partial}{\partial t} \Psi\bar{\Psi} = \Psi\Delta\bar{\Psi} - \bar{\Psi}\Delta\Psi = \text{div} (\bar{\Psi} \text{ grad } \Psi - \Psi \text{ grad } \bar{\Psi}). \quad (65)$$

The appearance of the gradient of a complex number may seem strange; but one may compute exactly as with the gradient of a real quantity, since it represents a combination of the gradient of the real part and that of  $i$  times the imaginary part.

Since  $\Psi$  has been normalized in such way as to make  $\int \Psi\bar{\Psi} d\tau = 1$ , the time rate of change of the charge density  $\rho$  is given by  $\partial(e\Psi\bar{\Psi})/\partial t$ . Also, as for any current, the equation of continuity of the electrical current is (cf. p. 196)

$$\frac{\partial\rho}{\partial t} = \text{div } \mathbf{i}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (66)$$

where  $\mathbf{i}$  is the current density vector. Thus the current density is represented in wave-mechanical form as

$$\mathbf{i} = \frac{eh}{4\pi im} (\bar{\Psi} \text{ grad } \Psi - \Psi \text{ grad } \bar{\Psi}). \quad . \quad . \quad . \quad (67)$$

The total current carried is given by

$$\mathbf{I} = \frac{eh}{4\pi im} \int (\bar{\Psi} \text{ grad } \Psi - \Psi \text{ grad } \bar{\Psi}) d\tau. \quad . \quad . \quad (68)$$

While this cannot be done for (67), it is possible to put (68) into a simpler form. Inasmuch as  $\Psi$  vanishes exponentially at infinity, the integral  $\oint \Psi \bar{\Psi} d\tau$  over an infinite sphere is zero. Hence by equation (59), p. 31,

$$0 = \oint \Psi \bar{\Psi} d\mathbf{s} = \int \text{grad } \Psi \bar{\Psi} d\tau = \int (\Psi \text{ grad } \bar{\Psi} + \bar{\Psi} \text{ grad } \Psi) d\tau,$$

and so the total current becomes

$$\mathbf{I} = \frac{eh}{2\pi im} \int \bar{\Psi} \text{ grad } \Psi d\tau. \quad . \quad . \quad . \quad . \quad (69)$$

Since we have  $\Psi = \psi e^{i\omega t}$  and  $\bar{\Psi} = \bar{\psi} e^{-i\omega t}$ , it is also possible to write

$$\mathbf{I} = \frac{eh}{4\pi im} \int \bar{\psi} \text{ grad } \psi d\tau. \quad . \quad . \quad . \quad . \quad (69')$$

If  $\Psi = \psi e^{i\omega t}$  and if  $\psi$  is real, the expression for  $\mathbf{i}$  vanishes. This may at first sight seem surprising, for on p. 698 we used real  $\psi$  functions in treating the hydrogen atom. However, we know from its magnetic behaviour that a rotational current is present in this case, as required even by the revolving electron of the simple atom model. In reality, the existence of a magnetic field compels us to carry out the hydrogen calculation using complex  $\psi$  functions, which yield a current. According to wave mechanics there can be no current without a magnetic field, since the  $\psi$  functions give the mean values, and in the absence of a magnetic field all orientations of the orbits occur with equal frequency and so cancel on the average.

### (c) *Electron in a Periodic Potential Field*

Let us again restrict our consideration to the one-dimensional case, which depicts all the essential facts concerning the behaviour of electrons in a space lattice. The form of the Schrödinger equation that does not involve the time is

$$\frac{d^2\psi}{dx^2} + \frac{8\pi^2m}{h^2} (E - V) \psi = 0, \quad . \quad . \quad . \quad . \quad (70)$$

where  $V$  is assumed only to vary periodically in space with an interval equal to the lattice constant  $a$ , i.e.

$$V(x + a) = V(x). \quad . \quad . \quad . \quad . \quad (71)$$

If  $\psi_m(x)$  is a solution corresponding to the proper value  $E_m$ , then  $\psi_m(x + a)$  is another solution corresponding to the same characteristic value; because if we set  $x + a = x'$  in (70) we get

$$\frac{d^2\psi}{dx^2} + \frac{8\pi^2m}{h^2} [E - V(x' - a)] \psi = 0. \quad . \quad . \quad . \quad (72)$$



On account of the periodicity of  $V$ , this is again identical with (70). Since we rule out the degenerate case, this is possible only if both solutions differ by a constant factor:

$$\psi(x + a) = p\psi(x). \quad . . . . . (73)$$

In a crystal of some extent, the condition  $\psi$  must recur after a certain distance. Even if we take the whole crystal to be a unit for purposes of calculation, we may imagine it to be succeeded by an identical region. In the one-dimensional case, assume this fundamental unit region to be

$$L = Ga, \quad . . . . . (74)$$

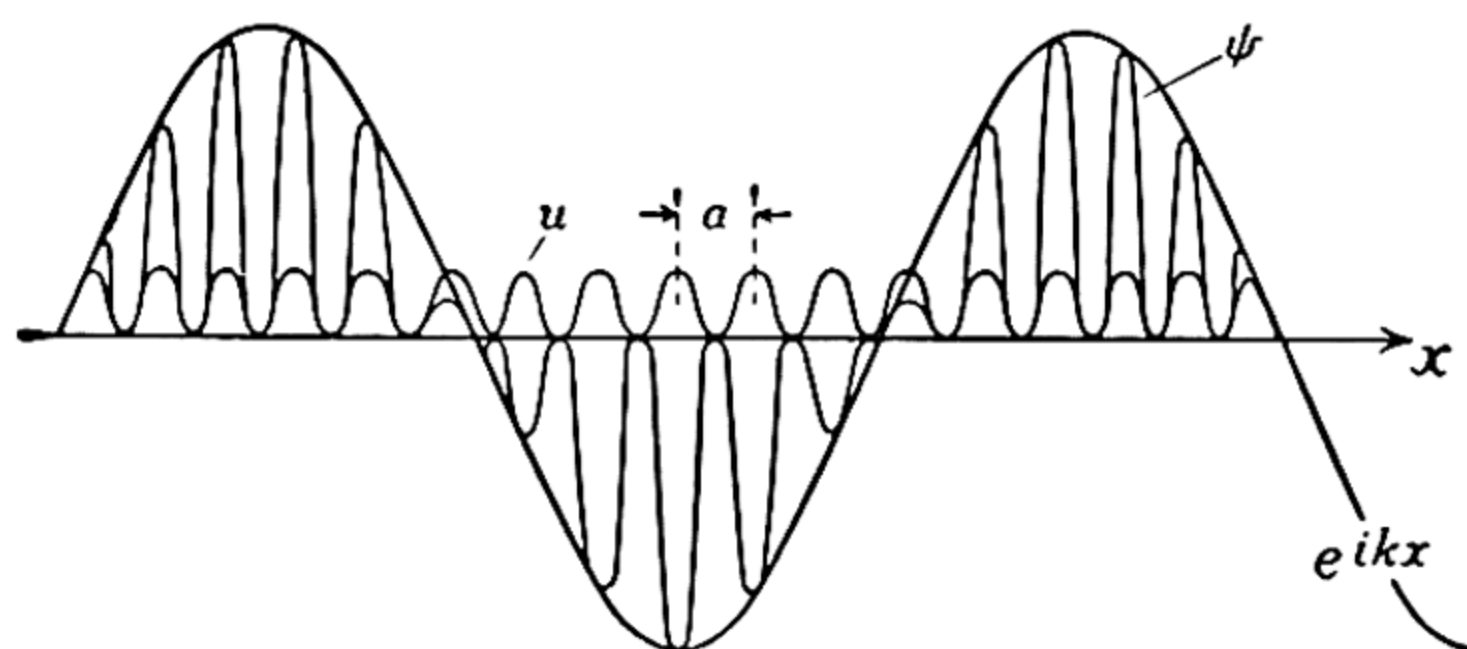


Fig. 5

where  $G$  is a very large number. Since the potential is due to the ions of the lattice,  $G$  also is the number of ions in the unit region. By repeated application of (73),

$$\psi(x + Ga) = \psi(x) = p^G \psi(x),$$

from which it follows that

$$p = e^{2\pi im/G} \text{ and } \psi(x + a) = e^{2\pi im/Ga} \psi(x), \quad m = 1, 2, \dots G. \quad (75)$$

This property is possessed by the function

$$\psi(x) = e^{2\pi imx/Ga} u(x), \quad . . . . . (76)$$

if  $u$  is a periodic function whose period is  $a$ . If  $V$  is zero, the solution is simply

$$\psi(x) = e^{2\pi imx/Ga} \times \text{const.} \quad . . . . . (77)$$

This represents a plane wave which moves from left to right or from right to left, depending on the algebraic sign. The periodic potential field, in conjunction with  $u(x)$ , results in a spatial modulation of the wave (cf. p. 57 and fig. 5 here).

Introduce  $k_m$  as an abbreviation for  $2\pi/\lambda$ . Then

$$k_m = \frac{2\pi m}{Ga}, \quad . . . . . (78)$$

whence

$$\lambda_m = \frac{Ga}{m}, \quad m = 1, 2, \dots G. \quad \dots \dots (79)$$

According to p. 695, the relation between the wave-length and speed of a free electron is

$$\lambda = \frac{h}{m_{el}v} = \frac{2\pi}{k}, \quad \text{or} \quad v = \frac{kh}{2\pi m_{el}}. \quad \dots \dots (80)$$

Consequently the energy of the free electron is

$$E = \frac{1}{2}m_{el}v^2 = \frac{k^2h^2}{8\pi^2m_{el}} = \frac{k^2}{\mu}, \quad \dots \dots (81)$$

where

$$\mu = \frac{8\pi^2m_{el}}{h^2}. \quad \dots \dots (82)$$

Since a free electron may have any speed,  $E$  and  $k$  are related parabolically.

As long as the potential is small compared with the electron energy, we can start from the solution for free electrons and proceed by the established methods of quantum-mechanical perturbation theory, considering the periodic potential as a small perturbation. For this purpose we develop  $V$  in a Fourier series:

$$V = \sum_{n=-\infty}^{n=+\infty} v_n e^{2\pi i n x/a}, \quad \dots \dots (83)$$

where, according to p. 55,  $v_{-n} = \bar{v}_n$ . Further, let  $V$  be normalized to make the mean value  $v_0$  equal zero.

The coefficients  $v_n$  are to be considered given, and we seek the Fourier coefficients, which replace  $u_m$  in (76):

$$\psi_m = e^{ikm x} u_m(x). \quad \dots \dots (84)$$

For simplicity, we omit the index  $m$  and write

$$u = \sum_{n=-\infty}^{n=+\infty} c_n e^{2\pi i n x/a}. \quad \dots \dots (85)$$

When these Fourier series are substituted in (70), the product of two series appears in the term  $V\psi$ . If the terms having the same exponent are made to vanish, equations determining the  $c_n$  are obtained. For instance, collecting the terms involving  $e^{2\pi i \cdot 4x/a}$ , the linear parts of (70) yield a term containing  $c_4$ , but the product  $V\psi$  yields a sum of

terms whose index sum is 4, e.g.  $v_1c_3$ ,  $v_2c_2$ ,  $v_3c_1$ . The equations, solved for  $c_n$ , are

$$c_n = \frac{\sum_p v_p c_{n-p}}{E - \frac{(k + 2\pi n/a)^2}{\mu}} \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad (86)$$

As long as the denominator does not vanish, all of the  $c_n$ , with the exception of  $c_0$ , may be allowed to approach zero as  $v_n$  does. This means that the proper functions and proper values of the electrons in the weak, periodic field differ but slightly from those in the absence of forces. However, it must be noted that the  $k_m$  now form a discrete, yet very dense, set. The situation is different, however, if the denominator vanishes; then the  $c_n$  can no longer be small. Thus for  $k = -\pi/a$  we have

$$c_1 = \frac{\sum_p v_p c_{1-p}}{E - \frac{\pi^2}{a^2\mu}}, \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad (87)$$

where, according to (81),  $E = \pi^2/a^2\mu$ .

The same denominator is obtained for  $c_0$ . In order to investigate the character of the result at these places we approach  $-\pi/a$  from neighbouring points by trying to determine  $c_0$  and  $c_1$  for  $k = (-\pi/a) + \delta$ , the corresponding proper value being  $E = (\pi^2/a^2\mu) + \epsilon$ .

Neglecting all coefficients shown above to be small, we get the following equations for  $c_0$  and  $c_1$ :

$$\left. \begin{aligned} c_0 \left( \epsilon + \frac{2\pi\delta}{a\mu} \right) &= v_{-1}c_1, \\ c_1 \left( \epsilon - \frac{2\pi\delta}{a\mu} \right) &= v_1c_0. \end{aligned} \right\} \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad (88)$$

These linear, homogeneous equations can be satisfied only by the vanishing of the determinant. This leads to an equation for  $\epsilon$ :

$$\epsilon = \pm \sqrt{v_{-1}v_1 + \frac{4\pi^2\delta^2}{a^2\mu^2}},$$

or, because  $v_{-1} = \bar{v}_1$ ,

$$\epsilon = \pm \sqrt{|v_1|^2 + \frac{4\pi^2\delta^2}{a^2\mu^2}}. \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad (89)$$

In the limit, with  $\delta = 0$ ,

$$\epsilon = \pm |v_1|. \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad (90)$$



Thus for  $k = -\pi/a$  we get two values of  $E_k$ , of which the smaller is a continuation of the  $E$  values that correspond to smaller values of  $k$ . But this means that the parabola mentioned on p. 729 has a break at  $k = -\pi/a$ , and the same thing occurs at  $k = \pm 2\pi/a, 3\pi/a$ , &c. Thus there are breaks in the spectrum of proper values. These places correspond exactly to the reflection in depth of the de Broglie waves from the row of points, similar to that of X-rays considered on p. 388. For  $k = \pm n\pi/a$  the waves reflected from the several points are in phase, so that no wave moving from left to right can arise. In the three-dimensional case these breaks occur for those directions of propagation

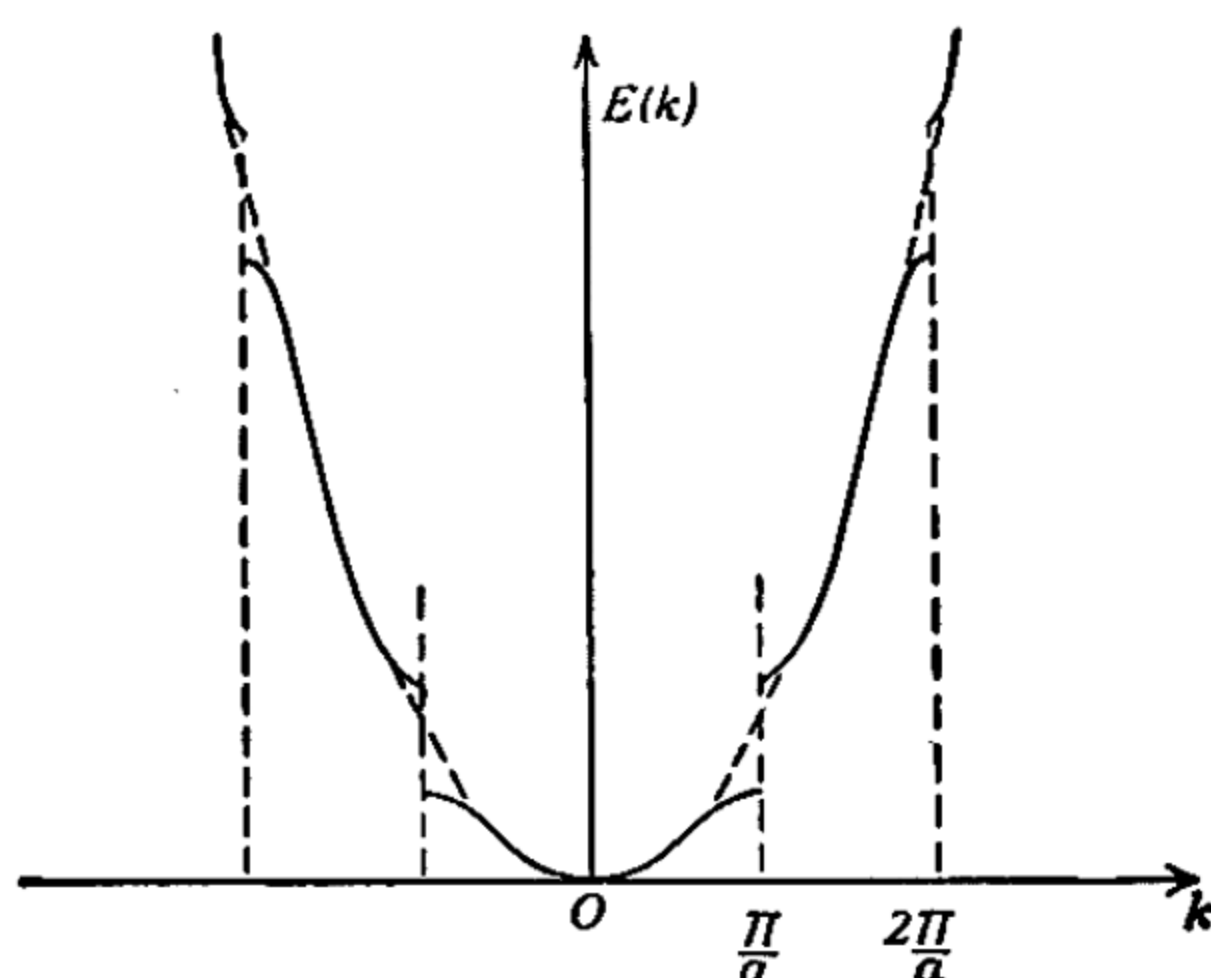


Fig. 6

and those wave-lengths for which, according to p. 388, reflection takes place. The trend of the proper values is as shown in fig. 6.

It is important, for what follows, to notice that the curve is horizontal at the points of discontinuity, as computation shows.

With increasing perturbation the breaks in the energy spectrum become broader, since the jump in  $E$  at  $k = n\pi/a$  is conditioned by  $v_n$ . However, for large perturbation, the present point of departure of the calculation is no longer the correct one. In that case, one starts with the proper functions of the electrons bound to an ion of the lattice. The result is that for very strongly bound electrons (K shell) the influence of the neighbours ceases to be important. For the outer electrons the sharp term values of the free atom change to broad *bands* that correspond to the above, except that the breaks no longer occur exactly at  $n\pi/a$ . These bands are designated according to the terms of the free atoms from which they developed. Thus one speaks of the 4s or of the 3d band of potassium.

(d) *Relation between Current and Energy Values*

In order to connect the behaviour of electrons in a periodic potential field with the conduction properties of matter we shall deduce an important relation between the current and the function  $E(k)$ . Despite the fact that the proper values constitute a discrete set, they lie so close together that they may be considered continuous inside a band  $E$ . If  $k$  is assumed to increase by a small amount  $\alpha$ , the  $E_k$  will change by an amount  $\beta$  and the function  $u_k$  (which has a period  $a$ ) will change by a function  $v$ . Thus to  $k + \alpha$  there corresponds the proper value  $E_k + \beta$  and the proper function  $e^{i(k+\alpha)x}(u_k + v)$ , or, putting

$$e^{ik}v = \chi, \quad . . . . . (91)$$

we have

$$\psi = (\psi_k + \chi)e^{i\alpha x}. \quad . . . . . (92)$$

To determine  $\beta$  and  $\chi$  we substitute into the Schrödinger equation. Since  $\psi_k$  satisfies equation (70), we get the following equation for  $\chi$  by discarding the products  $\chi\alpha$  and  $u\beta$ :

$$\frac{d^2\chi}{dx^2} + \mu(E_k - V)\chi = -\beta\mu\psi_k - 2i\alpha\frac{d\psi_k}{dx}. \quad . . (93)$$

According to p. 708 this inhomogeneous equation will have a solution only if the solution  $\chi = \psi_k$  of the homogeneous equation is orthogonal to the right side. The situation differs from that on p. 708, however, in that  $\psi_k$  is complex. In this case the condition for orthogonality is similar to that on p. 708 except that in place of  $\psi_k$  we have the conjugate value  $\bar{\psi}_k$ . Then, since  $\int \bar{\psi}_k \psi_k d\tau = 1$ ,

$$\beta = -\frac{2i\alpha}{\mu} \int \bar{\psi}_k \frac{d\psi_k}{dx} d\tau. \quad . . . . . (94)$$

By equating (69), p. 727, the current is given by

$$\mathbf{I} = \frac{eh}{2\pi im_{el}} \int \bar{\Psi}_k \text{grad } \Psi_k d\tau = \frac{eh}{2\pi im_{el}} \int \bar{\psi}_k \text{grad } \psi_k d\tau. \quad (95)$$

Also, in the one-dimensional case,

$$\text{grad } \psi_k = \frac{d\psi_k}{dx}.$$

Then, on account of (82), p. 729, we have the important relation

$$\frac{p}{\alpha} = \frac{dE_k}{dk} = \frac{h}{2\pi e} I, \quad . . . . . (96)$$

or

$$I = \frac{2\pi e}{h} \frac{dE}{dk}. \quad . . . . . (97)$$

The quantity  $dE/dk$  vanishes at the edges of the bands, as mentioned above. Hence the current, which is carried by electrons having the limiting energy, is zero. Physically, this remarkable result means that because of the reflection in depth of the de Broglie waves there is now a standing wave rather than a progressive one.

*(e) Conduction Properties of Matter; Algebraic Sign  
of the Hall Effect*

We are now in possession of all the essentials for building the wave-mechanical theory of metals. To be added, as in the case of an electron gas, is Pauli's principle, i.e. the Fermi Statistics. We are then in a position to give qualitative answers to the more important questions. Space does not allow a description of the quantitative three-dimensional theory, for which the special literature should be consulted.\*

A band contains  $G$  closely spaced energy levels. The Pauli principle requires that every quantum state be only singly occupied, and on account of the two possible spin orientations we can thus accommodate  $2G$  electrons in each band. There are  $G$  atoms in the unit region. Without regard to spin, each electron in an atomic configuration is characterized by the quantum numbers  $n, l, m_l$  (cf. p. 679). For each trio of numbers, designating a single band, there are two possible electrons per atom. This is precisely the case for closed shells. For example, this is true for the inner as well as for the outermost shells of the alkali halide ions, which are like the rare gases in this respect. The alkalis, on the other hand, have a single electron outside the rare-gas configuration. But, because of the spin, the Pauli principle allows for the presence of two such electrons, so that the band corresponding to the valence electron is only half-occupied.

What influence on conductivity does the extent of occupation of the bands have? In an electric field the electron obtains energy and so moves up to the higher levels within a band. In general, however, the energy that can be obtained from the field is not enough to force a transition to the next band. If the upper levels are already occupied (rare gases) a displacement of this kind is not possible, and the material is an insulator. The case is otherwise for the alkali atoms, where the half-filled state of the bands provides enough room for such displacements, and so the alkalis in the solid state are typical examples of metallic conductors.

The case of the alkaline earths is more complicated. The  $ns$  shell is fully occupied by two electrons. That these elements are nevertheless conductors is explained because the adjoining band lies so close to the  $ns$  band that the two overlap. The effect of the field is a double

\* See, for example, F. Seitz, *The Modern Theory of Solids*, McGraw-Hill Book Co., 1940.



one: the number of electrons in the next higher band is increased, and vacancies occur in the  $ns$  band. The effect of these vacancies will be explained below. Another situation that may arise is that two bands are separated, but that the distance between them is so small that thermal motion suffices to elevate individual electrons to the upper band. Substances where this occurs are the simplest electronic semi-conductors. For such materials the conductivity does not decrease with increasing temperatures, as it does for metals; rather, corresponding to the increasing number of electrons reaching the next band, the conductivity increases exponentially.

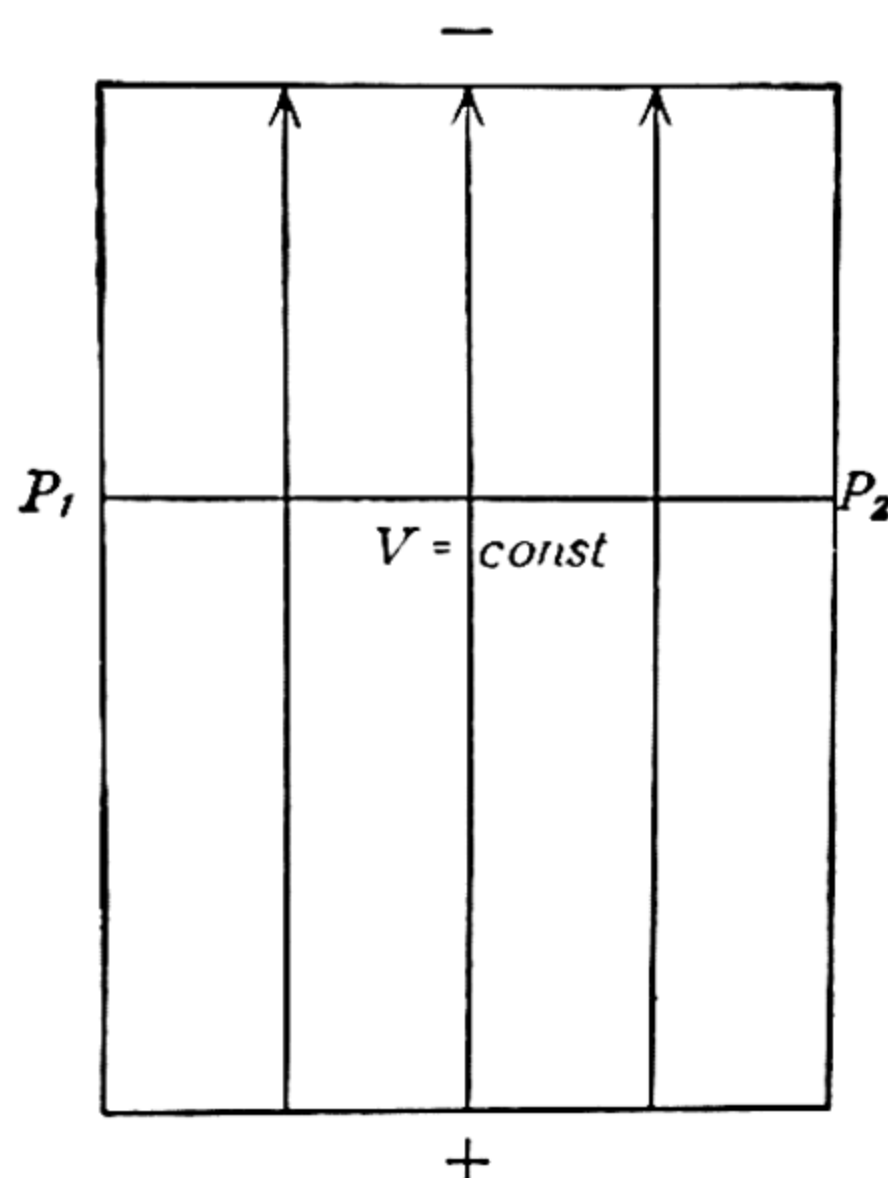


Fig. 7

The Hall effect, mentioned above on several occasions, is the following phenomenon: Let a current flow through a rectangular plate in the direction of its longer side (fig. 7). The equipotential lines will be perpendicular to the lines of flow. Between two points  $P_1$  and  $P_2$  lying on an equipotential there will be no potential difference, and a voltmeter joining them will read zero. If we now apply a magnetic field directed outward and perpendicular to the page, there will be a force  $-e[\mathbf{v}\mathbf{H}]$  on each moving electron and it will be deviated to the right.\* The accumulation of electrons on the right side gives rise to a transverse electric field held in equilibrium by the magnetic force. The next effect is that the lines of flow remain parallel to the side of the sample, but the points  $P_1$  and  $P_2$  no longer are at the same potential. For electrons, which are carriers of negative charge,  $P_2$  must be negative with respect to  $P_1$ . This is the case, for instance, for alkalis.

\* Notice that  $\mathbf{v}$  for an electron is directed from  $-$  to  $+$ .

For many metals, such as zinc, the sign is reversed, as if positive carriers were taking part in the flow of current. The band theory accounts for this readily: It seems fairly obvious—and may be proved strictly—that the “hole” in an almost fully occupied band acts like a positive charge. In an instance where two bands overlap as mentioned above, which happens for the elements of the second column of the periodic table, the conduction is made up of two parts—that of the holes and that of the electrons of the higher band. If the former preponderates, the sign of the Hall effect reverses.

The next task is to compute the conductivity and its temperature dependence on the basis of our picture of electron waves permeating the metallic crystal. As long as the lattice is a perfect one there will be no possibility, according to the above treatment, of transferring energy to the atoms of the lattice, i.e. a uniform metal has zero resistance at absolute zero. At higher temperatures, however, the periodicity is disturbed by thermal motions and the randomly distributed departures may be introduced into the computation as a perturbation. The calculation is cumbersome but it succeeds in giving a satisfactory picture of conduction.

## 6. The Role of Lattice Defects in Dielectric Crystals.

It was stated in the preceding section that a material is an insulator or a conductor according as the uppermost energy band is completely or partially occupied. In rare cases, the nearest unoccupied band may lie such a short distance above the highest fully-occupied one that even at room temperature occasional electrons may find their way into the upper band. Each electron thus elevated contributes in two ways to the conduction: firstly, because this electron is itself mobile, and secondly, the hole now remaining in the previously fully-occupied band is able to move (“hole conduction”). Since the number of electrons elevated in this way is governed by the Boltzmann distribution, the conductivity of such truly electronic semiconductors increases with the temperature according to  $e^{-W/kT}$ , where  $W$  is the excitation energy. This temperature dependency is well founded experimentally.

With other substances, e.g. boron, the conductivity of different samples is often found to vary by several orders of magnitude, even at the same temperature. The cause, recognized only relatively recently, is the great influence of lattice defects or imperfections. These may be atoms of impurities, holes, or atoms of the material itself which occupy locations between lattice points. If they are present in concentrations as low as even  $10^{-5}$  they are able to alter the conduction properties completely. The lattice defects constitute the critical factor in the technical application of a tremendous number of materials: oxide cathodes of electron tubes, dry rectifiers, detectors, transistors, selenium photocells, phosphors, and photographic emulsions. Preparation of



suitable materials for such uses amounts to finding ways of introducing the proper defects into the substance. We shall consider two classes of materials here—phosphors and photographic emulsions.

One of the oldest and most widely used phosphors is  $\text{ZnS}$  contaminated with traces of other heavy metals or rare earths. It may be fabricated in forms that show long-lasting phosphorescence, or others where the decline is rapid, such as those used to coat cathode-ray and television receiving tubes. In the level scheme of such a phosphor there is an impurity band  $C$ , of finite width, lying higher than the fully-occupied band  $A$  of the main material (fig. 8). An electron at the location of the impurity ion has somewhat higher energy than one in the main lattice. In every phosphor there is a second kind of defect fulfilling the requirement that the whole crystal be electrically neutral. If, for example, the place of a  $\text{Zn}^{++}$  ion is taken by a  $\text{Cu}^+$ , there will be an excess of negative charge in the neighbourhood. This may be compensated, for instance, by the removal of one  $\text{S}^{--}$  ion for each two  $\text{Cu}^+$  ions present. Since a vacancy of this kind corresponds to a positive charge, it will attract and hold electrons, acting like an electron trap. The corresponding energy level  $D$  lies just below the conduction band  $B$ .

The mechanism of phosphorescence is then visualized as follows: A photon absorbed by the main substance ejects an electron from  $A$  and sends it upward to  $B$ .\* In  $B$  the freed electron interacts with the lattice elements, losing energy thereby and descending to the lower edge of band  $B$ . Its return from this point to  $A$ , which would violate a selection rule, is prevented by the circumstance that its former place in  $A$  was soon filled by an electron from the impurity level. A lattice defect in the latter level is completely surrounded by normal lattice elements, and so such a transition can take place readily.

The conduction electron, whose existence can be directly verified electrically, moves through the lattice and is trapped in  $D$ . Passage from here to a vacant place in  $C$  will be possible only if the level  $D$  happens to lie near an impurity atom in which there is a vacancy. This seldom happens, and the electron usually is thrown back into the conduction band by the absorption of thermal energy. This happens repeatedly until the electron chances to fall into a properly placed trap from which it can return to an available place in an impurity ion. The latter transition is accompanied by the emission of radiation, and the phosphor is then back in its original condition.

This picture of what goes on, while apparently somewhat artificial, succeeds in bringing order to a wealth of observational material. For one thing, it explains at once the strong temperature dependence of the luminescence which enables the substance to “freeze” the absorbed light and then give it out again rapidly upon warming. In order to

\* The rarer transitions  $C$  to  $B$  and  $C$  to  $D$  occur also, but will be neglected here, since there are many more absorbing atoms in the ground level.



pass from an unfavourably placed trap to a better one located near an impurity ion that is able to accept an electron, the conduction electron must take up heat energy to bring it back to the conduction band. The band theory also explains the selective effect of infra-red light whose quanta correspond to the energy difference  $B - D$ . The mechanism is simpler only for certain solid organic phosphors, where the absorption and emission processes take place within the same molecule, the excited state being metastable.

There is a connexion between the explanation of the luminescence of inorganic phosphors and that of the latent photographic image. The energy level scheme corresponds to that of fig. 8. The effective impurity centres are known to be  $S^{--}$  ions, which find their way into the silver bromide granules from the gelatin of the emulsion. When the necessity of sulphur impurities was finally discovered, many puzzling cases of insensitivity in emulsions were cleared up.

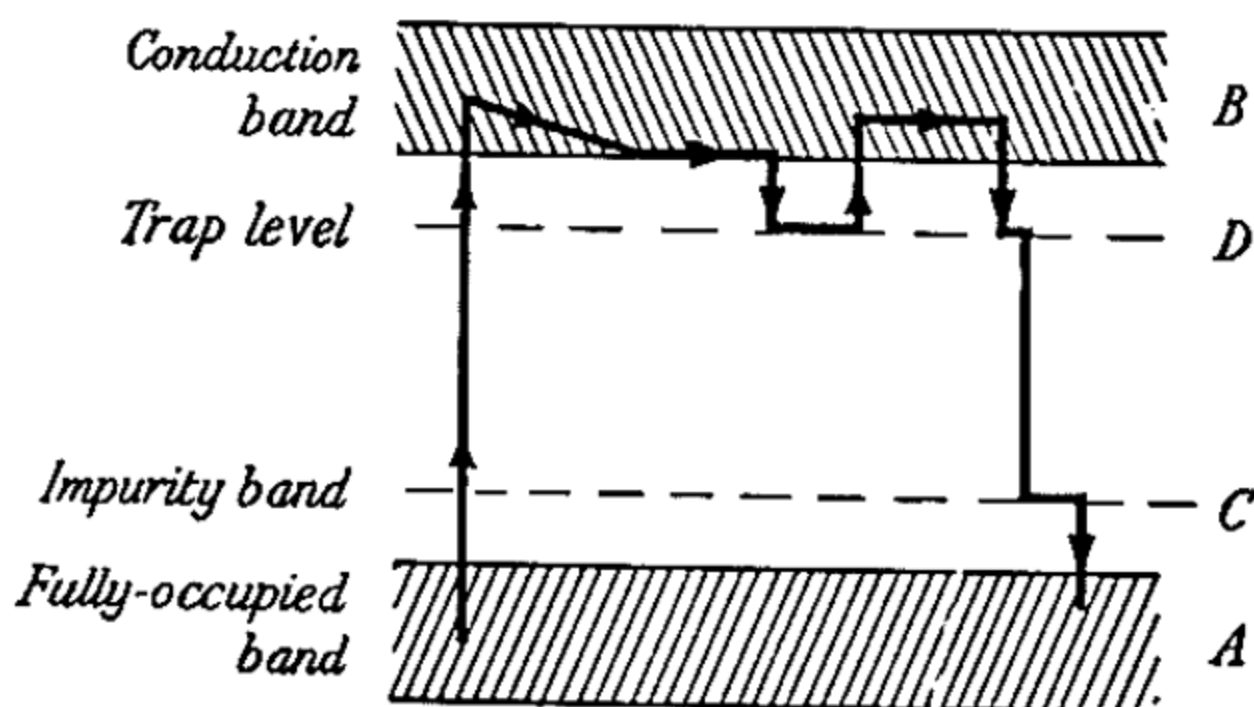


Fig. 8

In order to compensate the double charge of each  $S^{--}$  ion, there must somewhere be a missing  $Br^-$  ion. These  $Br^-$  vacancies are the electron traps. Compared with phosphors, the new feature in addition to the purely electronic process is this: passage of an electron to a trap site is followed by a motion of ions, which leads to the formation of sensitivity centres. Even at room temperature, the silver halides exhibit an appreciable electrolytic conduction, borne mainly by the silver ions. This has to do with the smallness of these ions. In the case of a photographic emulsion, there is also a shift of the  $Br^-$  ions to be taken into account, and there is a corresponding movement of  $Br^-$  holes. Such motion can take place if there is a negative space charge at some point. This will happen, for example, if a second electron should fall into the  $Br^-$  hole. Experiments with alkali halides bear this out. In such instances, little clusters consisting of  $Ag^+$  ions and  $Br^-$  holes containing one or two electrons are formed. Except for the size of the lattice constant, such an aggregate is nothing more nor less than a tiny crystal of metallic silver, which serves as a sensitivity centre.

The decrease in sensitivity at low temperatures is then understandable on the ground of reduced ionic mobility. Further, the quenching of the latent image by infra-red light (Herschel effect) is explained as the elevation of the trapped electron to the conduction band. Also, the fact that the photographic density is found to depend not only on the product of intensity by time but also on the way in which the energy is introduced during exposure is explained by the above considerations.

## CHAPTER XLII

### NUCLEAR PHYSICS

#### 1. General Remarks on the Physics of the Atomic Nucleus.

Now that the structure of the electronic shells of the atom and the processes going on in them have been revealed in their main outlines, the interest of physicists since about 1930 has chiefly been centred on the atomic nucleus. Here both experiment and theory indicate relationships essentially different from those which apply to the outer layers of electrons, and we shall accordingly begin by discussing these differences.

The most important distinctive feature of the physics of the nucleus is the immense amount of energy which, as has long been known from the facts of radioactivity, is involved in any nuclear phenomenon. While the amounts of energy involved in changes in the outermost shell of electrons (excitation, ionization, chemical binding) are of the order 5 electron volts, the amounts of energy transformed in nuclear processes are over  $10^6$  times as great. This energy ratio is even greater than we should expect from the dimensions of the nucleus and of the outer shells—that is, by comparing the potential energy of two elementary charges, first when at a distance apart of the order of the diameter of the atom ( $10^{-8}$  cm.) and again when separated by a distance of the order of size of the diameter of the nucleus ( $10^{-13}$  to  $10^{-12}$  cm.). Owing to the magnitude of the energy liberated it is possible, in contradistinction to the physics of the outer shells, to observe *individual* phenomena. In this way we are made aware of uncommon occurrences, which, though they elude the methods of large-scale physics, may have an extremely important bearing on cosmological problems, because there the rarity of the event is compensated for by the immense periods of time available. These large amounts of energy manifest themselves in the form either of fast corpuscular rays (electrons, protons and  $\alpha$ -rays) or of radiation of very short wave-length ( $\gamma$ -rays). In the case of the short  $\gamma$ -rays, however, it must be borne in mind that though we speak of *waves* practically nothing of a wave nature is revealed experimentally, all “wave-length determinations” being really determinations of the magnitude of the light quantum  $h\nu$ . Conversely, these energy-rich corpuscular rays and  $\gamma$ -rays may be



used to bring about changes in the nucleus. We have them at our disposal in the rays emitted by the radioactive elements and in the cosmic rays. The latter were discovered by V. F. Hess as far back as 1910 and have since been the subject of an enormous number of experiments. It may be taken as established that these rays are diffusely incident on the earth from without, coming from all directions in space, and that they are a mixture of the most diverse types of ray, whose energy may reach  $10^{17}$  electron volts and perhaps even more.

In recent years we have found means of giving protons, deuterons and  $\alpha$ -particles velocities far in excess of those of "natural" radiations (except those of cosmic-ray particles). Thus, in the cyclotron  $\alpha$ -particles may be given energies amounting to several thousand million electron volts, as compared with  $\alpha$ -rays from radioactive substances, whose energy is never as much as ten million electron volts. Further, particularly energetic particles are obtained also from artificially radioactive elements (cf. § 4).

Several methods are available for detecting the individual phenomena. The simplest of these is the scintillation method, which is very suitable for  $\alpha$ -particles and hydrogen nuclei or protons. The energy of these is so great that when an individual particle hits a screen of zinc blende it gives rise to a flash of light which is visible in a microscope. Owing to the differing brightness of the flashes it is possible to distinguish between  $\alpha$ -particles and the feebler protons. The velocity of the particles is determined from their range, i.e. the distance from the source of radiation, in air at  $0^\circ$  C. and 760 mm. pressure, at which they can no longer produce scintillation. The relationship is an empirical one; for  $\alpha$ -particles the formula

$$R = 0.965 \times 10^{-27} v^3, \quad . . . . . (1)$$

where  $R$  is the range and  $v$  the velocity, is widely applicable. According to Blackett's experiments, in which the velocity was obtained from the magnetic deviation, the formula

$$R = 1.77 \times 10^{-31} v^{3.4} \quad . . . . . (2)$$

is more satisfactory for protons. A formula which has a more accurate theoretical basis, but which is very inconvenient in practice, has been given by Bethe. Reference may be made to equation (5) on p. 742.

Still more elegant and direct is the cloud chamber method due to C. T. R. Wilson. In this method the paths of the particles are made visible owing to the ions formed along these paths being caused to act as nuclei for the condensation of supersaturated water vapour. Here again the nature of the particles is revealed simply by the intensity of the track (the paths of electrons can also be made visible in the same way). As before, the velocity is obtained from the range, that is,

the length of the track. A better method consists in placing the chamber in a magnetic field and measuring the resultant curvature of the track. This method is especially useful in revealing the sign of the charge carried by the particle. Recently it has been found that  $\alpha$ -particles and protons leave tracks in a photographic emulsion—a method that is valuable for recording rare events without requiring special attention.

Again, the ions produced by a particle can be detected directly by means of specially sensitive electrometers, such as those due to Hoffmann. Or the ions first formed may be made to give rise by collision to a large number of new ions, so that less sensitive instruments may then be used to detect the total current. This is the principle of the Geiger and Geiger-Müller counting apparatus. In both types of instrument the ions first produced give rise to a gaseous discharge which is then rapidly extinguished by a suitable device. Depending upon the magnitude of the applied potential, a counter may be made to give a current pulse proportional to the number of initially produced ions or else each particle may be made to produce a pulse of the same strength. It is necessary to use weakly ionizing particles (electrons) in the breakdown range. Counting and recording devices may be actuated by means of suitable amplifiers. Recently, in place of some of the above instruments, scintillation counters have come into use. Here the flashes are amplified by photo-electric means and then registered. Suitable scintillation materials permit the recording of  $\beta$ - and  $\gamma$ -rays as well.

For very energetic particles such as those found in cosmic rays (§ 15) it is often difficult to distinguish their nature. The charge is usually recognized as a multiple  $z$  of the electron charge, but the rest mass and the velocity (or the kinetic energy) are unknown. Although the requirements are usually prohibitive, it may be possible to observe the path of the particle in a strong magnetic field directed normal to its path. In such case, equation (14) of p. 429 furnishes *one* relation. The product  $H\rho$ , called the magnetic rigidity, tells us the value of  $mvc/ze$ . But at the extremely high velocities encountered, the relativistic mass increase becomes important, and the value sought is the rest mass  $m_0$ , which may be very much smaller than  $m$ .

Properly speaking, the product  $H\rho$  gives us a relation between  $m_0$  and the kinetic energy  $E$ . Thus in equation (14), referred to above,  $m$  is to be replaced by  $m_0/\sqrt{1 - v^2/c^2}$ . From equation (82) on p. 257,

$$E = m_0 c^2 \left( \frac{1}{\sqrt{1 - v^2/c^2}} - 1 \right). \quad \dots \quad (3)$$

This may be solved for  $v$ . Using the result and also  $m_0/\sqrt{1 - v^2/c^2}$ , we find that

$$H\rho = \frac{1}{ze} \sqrt{E^2 + 2m_0 E c^2}. \quad \dots \quad (4)$$



A second relation is found from the energy loss per unit path length which the particle experiences by ionization of the air or other material through which it travels. Bethe derived a relation for very high energies  $E \gg eV_i$ , where  $V_i$  is the complete ionization potential of the atoms composing the medium; it may range as high as 10 kv. His equation for the energy loss per unit distance is

$$-\frac{dE}{dx} = \frac{4\pi e^4 z^2 N}{m_e v^2} \left\{ Z \left[ \log \left( \frac{2m_e v^2}{\bar{I} e_i} \right) - \log \left( 1 - \frac{v^2}{c^2} \right) - \frac{v^2}{c^2} \right] - C \right\}. \quad (5)$$

The mass of the particle does not appear, so that all particles having the same charge and speed will suffer the same loss. The charge number of the particle is  $z$ , while the atomic number of the atoms of the medium is  $Z$  and their mean ionization potential is  $\bar{V}_i$ . The number of atoms per unit volume of the medium is  $N$  and  $C$  is a correction term which may be neglected at high velocities.

Inspection of the equation discloses that the energy loss, and hence the intensity of ionization, decreases rapidly as the particle speed increases, approaching a flat minimum. For protons in air, this minimum corresponds to energies of about 2300 Mev. The number of droplets per unit length in a cloud chamber or the number of developed grains per unit length on a photographic plate increases with the number of ion pairs, so that the energy loss can be found empirically from such observations. The thickness and curvature of the tracks in a magnetic field almost always permit of unique identification. In the absence of a magnetic field, the energy may be found from the range of the particles. Suppose the energy loss has been determined as a function of the distance traversed. Numerical integration will then give  $E$  as a function of the distance  $x$  and hence  $-dE/dx$  as a function of  $E$ . Then the range is given by

$$R = \int_0^R dx = - \int_{E_0}^0 \frac{dx}{dE} dE. \quad (6)$$

Since  $R$  is observed, this is an equation for the determination of the unknown initial energy  $E_0$ . Because of the uncertainty of the energy loss measurements, one usually starts from the empirical range-energy relation for 15 Mev particles and integrates from this point up to the observed range. As a result of the simple dependence on  $z$  and on the particle mass  $M$ , equation (6) is especially suitable for computing the range of a particle of mass  $M'$  and charge  $z'$  from the known range of a particle of mass  $M$  and charge  $z$  having the same velocity.\* Taking the energy, as usual, to be the abscissa, corresponding points will have different abscissas as well as different ordinates. (Cf. *Ex.* 134 opposite.)

\*  $dE/dx$  is independent of  $M$ , but  $E$ , and therefore  $dE$ , is proportional to  $M$ .



In the case of electrons, there are two other mechanisms of energy loss in addition to that due to ionization: Compton effect and Bremsstrahlung (according to even the classical theory, an accelerated electron or other charged particle radiates electromagnetic waves—the continuous X-ray spectrum). Further, the formula for  $dE/dx$  includes other terms that can no longer be neglected when dealing with electrons, so that the range formula (5) is no longer applicable to electrons. For heavier particles the retardations are never so great as to make the Bremsstrahlung important.

So far as the *theory* of nuclear physics is concerned, the goal we set ourselves is the deduction, by application of wave mechanics, of the various nuclear states from some sort of model, as was accomplished in the case of the electron shells. Matters are still very much in a state of flux, but a number of important results have been obtained from the principles of conservation of momentum, energy and charge, whose validity is assumed to extend to the nucleus. Of quite special importance is the law of the inertia of energy (cf. p. 258), which takes the place of the conservation of matter. It enables us to calculate the release of energy in a nuclear reaction from the mass defect. In this connexion the following transformation formulæ should be noted:

$$1 \text{ Mev} = 1.60 \times 10^{-6} \text{ erg} = 1.78 \times 10^{-27} \text{ gm.}, \quad . \quad . \quad (7)$$

$$\text{Mass of electron} = 5.12 \times 10^5 \text{ electron volts.} \quad . \quad (8)$$

Thus if a mass decrease of  $10^{-3}$  a.m.u. (p. 765n) should occur in a nuclear reaction, energy amounting to about 1 Mev would be liberated.

*Ex. 134.* An  $\alpha$ -particle ( $M = 4$ ,  $z = 2$ ) with an energy of 12 Mev has an observed range of 14.42 cm. Find the energy of a proton ( $M = 1.008$ ,  $z = 1$ ) which will have this range.

## 2. The Hyperfine Structure of Spectral Lines as a Connecting Link between the Physics of the Outer Shells and that of the Nucleus.

In addition to the methods peculiar to nuclear physics which were described in the previous section, spectroscopy, which is a method of ordinary large-scale physics, has its contributions to make to the problems of nuclear structure. One way in which spectra are affected by the nucleus depends in the first instance on the joint motion of the nucleus and its surrounding shells; this, however, is only of importance in the lightest elements. In single-electron systems the motion about the common centre of mass can be calculated directly; we have already carried out this calculation on p. 652 for the interval between the H and  $\text{He}^+$  lines. This effect of the mass of the nucleus has gained fresh significance in connexion with the lines of the isotope of hydrogen of mass 2 discovered by Urey, Brickwedde, and Murphy, whose nucleus is called a *deuteron*. The lines of the isotope are found by replacing

$M_H$  in equation (31) (p. 652) by twice its value. The proportion in which the two isotopes occur may be deduced from the ratio of the intensities of their spectral lines. While ordinary hydrogen contains only one part in 5000 of the heavy isotope, it is easy to obtain a richer mixture, owing to the magnitude of the *relative* difference of mass. (With other elements, as is well known, the difficulties of separating isotopes in weighable quantity are very great.) One way of carrying out the separation, for example, is by the electrolysis of water, in which a preponderance of the lighter hydrogen isotope is set free.

Apart from this effect of the nucleus due purely to its mass, there is an electrodynamic interaction which gives rise to a splitting up of terms additional to that involved in the systematic theory of spectral terms developed on p. 669. This is referred to as *hyperfine structure*. The phenomenon has long been known in the case of the mercury lines; owing to the slowness of the splitting these lines have often served as tests for interferometers of high resolving power.

If we ascribe an angular momentum of magnitude  $Ih/2\pi$  to the nucleus it is possible to develop a systematic theory of the hyperfine structure. The total spin of the atom is then given by  $F$  (fine quantum number), the resultant of  $I$ , the moment of the nucleus, and  $J$  (inner quantum number), the moment of the electronic shells. The spin vectors  $I$  and  $J$  are combined to form the resultant  $F$  in the same way as  $L$  and  $S$  are combined to form  $J$ ,  $J$  corresponding to the vector  $L$ . When  $J > I$  a figure the same as that on p. 670 but with a different interpretation of the vectors gives a hyperfine structure with  $(2I + 1)$  sublevels to every term. The Zeeman effect may be treated in the same way. It is to be noted, however, that owing to the smallness of the splitting in fields of the necessary intensity there is an uncoupling of the vectors  $J$  and  $I$  corresponding to the Paschen-Back effect, which gives the Zeeman type of the multiplet component in question a hyperfine structure of  $(2I + 1)$  components. The mechanical moment of the nucleus is accordingly determined by the number of components. Very many observations of fine structure have been made and they show that *atoms of even atomic weight have integral, but in most cases zero, nuclear moment; whereas in atoms of odd atomic weight the nuclear moment is always equal to an odd multiple of one-half.*

The magnetic moment of the nucleus, however, is not determined by the mechanical moment of momentum, as we do not know whether the latter is due to circulating electrons or to nuclei. If we replace  $m$ , the mass of the electrons, in equation (60) (p. 666) by  $M_H$ , the mass of the proton, the magnitude so obtained is called a *nuclear magneton*; it is accordingly 1840 times as small as the Bohr magneton. If we make an approximate calculation of the term perturbation due to a magnetic dipole situated at the nucleus, the very small splitting gives the order of magnitude of the magnetic moment as that of a nuclear



magneton. Direct measurements, using molecular-beam techniques, lead to the same result. This furnishes one of the first objections to the earlier idea of nuclear electrons, whose spin would give rise to one Bohr magneton per electron.

From the discrepancy between the observed hyperfine structure and the separations to be expected from the multiplet rules, Schüller has drawn far-reaching conclusions about an ellipsoidal distribution of charge in the nuclei of many atoms (electric quadrupole moment).

In addition to this hyperfine structure occurring in a *single* isotope, there is another observed in elements which consist of a *mixture* of isotopes; here the lines or groups of lines correspond to the individual isotopes. The combination of these two effects make it difficult to analyse the hyperfine structure in elements such as mercury which have many isotopes. Inasmuch as the mass effect (p. 651) of the nucleus is least for the heavy elements, this type of hyperfine structure shows a difference in the nuclear field which is not that of a point charge. In fact, this isotope effect is greatest for the lowest *S* terms (e.g. the 6s), whose electrons approach the nucleus very closely owing to their large eccentricity.

### 3. Radioactivity.

Radioactive phenomena are amongst the earliest known facts of atomic physics. On them the establishment of an experimental basis for the model atom essentially depended. As is well known, the heaviest elements spontaneously emit rays which consist partly of radiation of short wave-length ( $\gamma$ -rays), partly of fast electrons ( $\beta$ -rays), and partly of  $\text{He}^{++}$  particles ( $\alpha$ -rays), the nature and energy of the rays being characteristic of the particular heavy atom in question. It was soon recognized that the source of these rays is the nucleus of the atom, so that the study of radioactivity represents the first phase of nuclear physics. As  $\alpha$ -rays and  $\beta$ -rays carry electric charges, the loss of one of these particles alters the charge on the nucleus and a new element is formed. Loss of an  $\alpha$ -particle diminishes the nuclear charge by two units, so that the element moves back two places in the periodic table.

Loss of an electron raises the positive nuclear charge by one unit, so that the new element is one place ahead of the former element in the periodic table. These are the celebrated displacement laws of Fajans and Soddy. If we trace out a disintegration series, we keep coming back to the same place in the periodic table, but with a different atomic weight each time. Scientists were thus forced to recognize the existence of isotopes, and these were found also among other elements by the methods of mass spectroscopy.

To describe the course of a transformation process as time goes on, we have to start from the experimental fact that the probability of any particular atom disintegrating in the following instant is quite



independent of external conditions, depending only on the internal structure of the nucleus. We accordingly represent this probability by a characteristic constant  $\lambda$ . The probability that a selected nucleus will disintegrate during the elementary interval of time  $dt$  is then  $\lambda dt$ , and if there are a large number ( $N$ ) of nuclei, the number of nuclei disintegrating in the interval  $dt$  is  $N\lambda dt$ , by the fundamental principles of statistics. We therefore have the differential equation

$$dN = -N\lambda dt \quad . \quad . \quad . \quad . \quad . \quad . \quad (9)$$

(the negative sign being necessary, as the number  $N$  is decreasing).

Integrating and taking  $N_0$  as the number of atoms when  $t = 0$ , we obtain the number of atoms left at the end of a finite time,

$$N = N_0 e^{-\lambda t}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (10)$$

That is, the disintegration constant  $\lambda$  is the reciprocal of the time required for the number of atoms of the original substance present to fall to  $1/e$  of its initial value. In addition the half-value period ( $T$ ) is often given, i.e. the time in which the number of atoms present is reduced to half its value. Substituting  $N_0/2$  for  $N$  in equation (10), we see that

$$T = \frac{1}{\lambda} \log_e 2. \quad . \quad . \quad . \quad . \quad . \quad . \quad (11)$$

As a rule the product of disintegration is itself radioactive. If, as in the case of uranium deposits, we have a long-lived parent substance whose diminution may be neglected, a state of equilibrium among the disintegration products is ultimately reached, atoms of each type being produced and being destroyed at the same rate. If we let the suffix 0 refer to the parent substance and the suffix 1 to the first disintegration product, the increase in  $N_1$  due to the disintegration of the parent substance is

$$dN_1 = N_0 \lambda_0 dt,$$

whereas the loss by further disintegration is

$$dN_1' = N_1 \lambda_1 dt.$$

In a state of equilibrium we accordingly have

$$N_0 \lambda_0 = N_1 \lambda_1, \quad . \quad . \quad . \quad . \quad . \quad . \quad (12)$$

or

$$\frac{N_0}{N_1} = \frac{T_0}{T_1}.$$

That is, *the numbers of atoms of the disintegration products of a long-lived parent substance present when radioactive equilibrium has been reached are in the ratio of the half-value periods.* This relationship serves to

determine the half-value periods of very long-lived types of atoms from those of other atoms which can be measured directly.

The fact that in the case of the  $\alpha$ -rays there is a definite connexion between the probability of disintegration and the energy of the particles shot out has long served as a stimulus to theoretical speculations on the mechanism of disintegration. For the shorter the life of an element, the greater is the energy set free in its disintegration. One might be tempted to compare this with the great velocity of strongly exothermic chemical reactions; here, however, the phenomena are quite different. The high velocity of a chemical reaction arises from the fact that the heat developed in the reaction increases the number and violence of the atomic collisions, whereas in radioactive transformations the relationship mentioned above applies to the *individual* atom. The relationship may be represented in a range extending over fifteen powers of ten by the following empirical formula involving two constants which are characteristic of the particular radioactive series:

$$\log \lambda = A + B \log E \quad (\text{Geiger-Nuttall}) \quad . \quad . \quad (13)$$

or

$$\log \lambda = a + b\sqrt{E}. \quad (\text{Swinne}). \quad . \quad . \quad (14)$$

The second form exhibits the rapid (exponential) growth of the probability of disintegration with the energy. This feature, however, is also brought out if we regard disintegration from the quantum mechanics point of view as the crossing of a potential barrier, such as we investigated in a very simple case on p. 721. For this phenomenon, which plays a large part in nuclear physics, we shall in future use the expression "quantum mechanics tunnel effect", as the particle, so to speak, gets past the potential hill by going through a tunnel.

Attempts have been made to find analogous formulæ for  $\beta$  disintegration, but here, as we shall see later (§ 9, p. 763), the theoretical aspects are more complicated. For artificially radioactive nuclei an empirical relationship (Sargent's formula) has been obtained which corresponds roughly to equation (13), except that the constant  $A$  has different values according as the nuclear momentum (nuclear spin) is altered by 0, 1 or 2 in the atomic transformation in question.

#### 4. Artificial Transformation and Excitation of the Nucleus.

The cohesion between the constituents of the nucleus is so great that we may expect *a priori* that it will be impossible to bring about any changes in the nucleus, unless by using corpuscular rays of very high energy. While electrons were until recently thought to have no effect, Rutherford succeeded as long ago as 1921 in bringing about a nuclear transformation by means of  $\alpha$ -rays, and proceeded to investigate the protons (hydrogen nuclei) liberated from hydrogen by  $\alpha$ -rays.



For the collision of  $\alpha$ -particle and proton the calculation carried out on p. 641 must be modified, inasmuch as the mass of the nucleus struck is not large compared with that of the  $\alpha$ -particle, but is actually smaller than the latter. There is no difficulty in calculating the motion of the two bodies relative to the common centre of mass. The velocity imparted to the hydrogen nucleus is very large; if the impact were exactly central it would be 1.6 times that of the  $\alpha$ -particle. Hence in the Wilson chamber the hydrogen nuclei behave in the same way as the  $\alpha$ -particles except that their tracks are somewhat finer. Their velocities are found to be distributed in accordance with theoretical expectations.

The really significant feature of this work of Rutherford's, however, was that he established the occurrence of these hydrogen nuclei or protons in nitrogen which previously contained no hydrogen whatever. Fig. 1 (Plate II) shows the liberation of a proton from nitrogen. The short track going sideways from the point of collision is that of the nucleus  ${}^8\text{O}^{17}$ , which consists of what is left after the nitrogen nucleus  ${}^7\text{N}^{14}$  and the  $\alpha$ -particle have united and discarded a proton.\* Thus the process consists not merely in the destruction of the nitrogen nucleus, but in the building-up of a new and heavier nucleus, a superfluous proton being ejected. Protons were subsequently found to be liberated from a very considerable number of elements, even the heavier ones, especially when very fast artificially produced  $\alpha$ -particles are used. In all these transformations, so far as Wilson chamber experiments show, the  $\alpha$ -particle remains in the nucleus.

The fact that in many cases the energy of the protons far surpasses that of the  $\alpha$ -particles used to liberate them (the range of the protons from aluminium is actually equivalent to a range in air of as much as 80 cm., the range of the original  $\alpha$ -particles being only 5 cm.) shows that what we are dealing with here is an elementary nuclear reaction, involving a transfer of energy ( $\Delta E$ ). This we shall reckon positive, as in thermochemistry, when energy is set free in the course of the transformation. If we attach the suffixes  $\alpha$ ,  $H$ ,  $k$  to all quantities referring to the  $\alpha$ -particle, the proton, and the newly-formed nucleus respectively, we have the energy equation.

$$\frac{1}{2}M_{\alpha}v_{\alpha}^2 + \Delta E = \frac{1}{2}M_H v_H^2 + \frac{1}{2}M_k v_k^2. \quad (15)$$

To express the conservation of momentum we denote the angles which the paths of the proton and of the nucleus make with the path

\* In nuclear physics it is customary to designate a given isotope by attaching the atomic weight as an index (upper right) and the nuclear charge as a suffix (lower left) to the chemical symbol. Further, a reaction is often written down in abbreviated form by putting first the bombarding particle and then the emitted particle in brackets between the initial nucleus and the final nucleus. Here, for example, we should write  ${}^7\text{N}^{14}(\alpha, p){}^8\text{O}^{17}$ . The symbols used are  $\alpha = \text{He}^{++}$ ,  $p = {}^1\text{H}^1$ ,  $d = {}^2\text{H}^2$ ,  $n = \text{neutron}$ ,  $e^- = \text{electron}$ ,  $e^+ = \text{positron}$ ,  $\gamma = \text{light quantum}$ ,  $\nu = \text{neutrino}$ ,  $m = \text{meson}$ .





Fig 1.—Ejection of a proton from nitrogen  
by an  $\alpha$ -particle (p. 748)  
[*Proc. Roy. Soc. A*, 136, Plate 8. fig. 1]



Fig. 2.—Disintegration of boron by protons (p. 751). Note  
the two distinct groups of  $\alpha$ -particles)  
[Photograph by F. Kirchner]



of the incident  $\alpha$ -particle by  $\phi$ ,  $\psi$  respectively; we then have the equations

$$M_H v_H \sin \phi - M_k v_k \sin \psi = 0, \quad (16)$$

$$M_H v_H \cos \phi + M_k v_k \cos \psi = M_\alpha v_\alpha. \quad (17)$$

From these we readily obtain

$$\frac{1}{2} M_k v_k^2 = \frac{1}{2 M_k} \{ M_H^2 v_H^2 - 2 M_H M_\alpha v_H v_\alpha \cos \phi + M_\alpha^2 v_\alpha^2 \}, \quad (18)$$

or, since  $M_H = 4 M_k$ ,

$$\frac{1}{2} M_k v_k^2 = \frac{M_H^2}{2 M_k} \{ v_H^2 - 8 v_H v_\alpha \cos \phi + 16 v_\alpha^2 \}. \quad (19)$$

Substituting this in (15) above, we obtain an equation which for a given value of the energy transfer gives  $v_H$  as a function of the angle  $\phi$  and the energy of the  $\alpha$ -particle. In many cases the energy transfer can be calculated from the mass defect and the distribution of  $v_H$  is in excellent agreement with theory. (The value of  $v_H$  is determined by the range, that of  $\phi$  from stereoscopic photographs.) In some cases several groups of protons were found, corresponding to several distinct values of  $\Delta E$ . This is obviously explained by the fact that when the energy of the protons is fairly small the nucleus is left in the excited state. The energy of excitation is probably given out later as a  $\gamma$ -quantum; the occurrence of  $\gamma$ -rays in nuclear disintegration has actually been established. All these observed facts justify the assumption of the validity of the energy and momentum theorems, as well as that of the existence of definite energy states of the nucleus.

The penetration of the  $\alpha$ -particle into the nucleus is obviously another case of the quantum mechanics tunnel effect. Here, however, the direction of motion is exactly opposite to that in radioactive disintegration, in connexion with which we first encountered the tunnel effect (p. 747). The  $\alpha$ -particle advances in opposition to the potential of the positively-charged nucleus, crosses the potential barrier, and reaches the interior of the nucleus. According to p. 747, the probability of crossing the barrier decreases rapidly as the difference between the energy required and the energy available increases. This is the reason why earlier attempts to shatter the nuclei of heavier elements with natural  $\alpha$ -particles did not succeed; in these, of course, the potential barrier is much higher, owing to the higher charge on the nucleus. On the other hand, the potential barrier, as was first recognized by Gamow, is lower for a proton than for an  $\alpha$ -particle, owing to the single charge of the former; this suggests the possibility of producing sufficiently fast protons to bring about atomic disintegration by means of potentials available in practice.



The first of these artificial transformations brought about without the use of radioactive substances was the decomposition of lithium by protons which had been subjected as hydrogen canal rays to a potential of about 100,000 volts, the experiments being made by Cockcroft and Walton working in Rutherford's laboratory. By means of the scintillations produced on a fluorescent screen they were able to detect the presence of  $\alpha$ -particles of energy  $8.6 \times 10^6$  electron volts. The behaviour of lithium, from which the yield is particularly great as compared with other atoms which can also be disintegrated by protons, was subsequently investigated very thoroughly. Wilson chamber photographs by Kirchner clearly showed that in each reaction two  $\alpha$ -particles appear. We accordingly have the equation

$${}_3\text{Li}^7 + {}_1\text{H}^1 + \frac{E(\text{H})}{c^2} = 2{}_2\text{He}^4 + 2 \frac{E(\alpha)}{c^2}, \quad . \quad . \quad (20)$$

or, in numbers,\*

$$7.0180 + 1.0081 + 0.0001 = 8.0077 + \frac{2E(\alpha)}{c^2}.$$

Using the convenient relation that 1 a.m.u. (atomic mass unit) is equivalent to 932 Mev (million electron volts), the above numbers show that an energy of 17.2 Mev has been released. The measured kinetic energy of both  $\alpha$ -particles exceeds that of the initial proton by about 17.0 Mev, in excellent agreement with the value computed by conservation of mass-energy.

The fact of the conservation of momentum is also established by Wilson cloud photographs. It is found that the two tracks of the  $\alpha$ -particles going in opposite directions enclose an angle of  $5^\circ$ , the value calculated for protons with the velocity used being  $4^\circ 50'$ . The discrepancy of  $10'$  is of course less than the experimental error.

Apart from these energy-rich  $\alpha$ -particles, the existence of two definitely slower groups of  $\alpha$ -particles has been established. These are the fragments arising from the disintegration of  ${}_3\text{Li}^6$  according to the reaction  ${}_3\text{Li}^6(p, \alpha){}_2\text{He}^3$ . That is, one particle is an " $\alpha$ -particle of mass 3". Hard  $\gamma$ -radiation is also observed, arising from the reaction  ${}_3\text{Li}^7(p, \gamma){}_3\text{Be}^8$ . The yield of disintegrations is represented by an exponential function, as one would expect in view of the occurrence of a tunnel effect. With the lowest potentials (20,000 volts)  $10^{15}$  protons

\* In nuclear reactions we are of course concerned only with the masses of the *nuclei*, but we actually use the masses of the atoms, i.e. of the nuclei plus the electrons, in our calculations. This does not introduce any error, since the number of electrons is the same on both sides of an equation representing a nuclear reaction. In problems of stability, in fact, we *must* carry out the calculations in this way (cf. Exercise 135, p. 763).

are required to bring about one disintegration, whereas with potentials of 200,000 volts only  $10^9$  protons are required.\*

The disintegration of boron has also been investigated very thoroughly. Here, in addition to  $\alpha$ -particles of certain definite ranges, there is a continuous velocity spectrum having a distinct maximum. This is perhaps due to the fact that the reaction takes place according to the equation

$${}_5\text{B}^{11} + {}_1\text{H}^1 + \frac{E(\text{H})}{c^2} = {}_3\text{He}^4 + \frac{3E(\alpha)}{c^2}, \quad . \quad . \quad (21)$$

or, in numbers,

$$11.0129 + 1.0081 + 0.0001 = 12.0117 + 3 \frac{E(\alpha)}{c^2}.$$

When, as here, there are *three* fragments the energy and momentum theorems are insufficient to determine the distribution among the individual particles. The reaction  ${}_5\text{B}^{11}(p, \alpha){}_4\text{Be}^8$  also takes place. This transformation gives rise to particles of definite range (see fig. 2, Plate II, facing p. 748), while the reaction of equation (21) corresponds to the shorter tracks seen in the picture.

Particular importance attaches to disintegration experiments with fast *deuterons* produced like the protons as hydrogen canal rays. When lithium is bombarded by a mixture of deuterons and protons a fresh group of  $\alpha$ -particles with the high energy of  $12 \times 10^6$  electron volts is produced, over and above the groups liberated by bombardment with protons only. On the basis of the mass defect this new group corresponds within the limits of experimental error to the disintegration of the lithium isotope 6 in accordance with the equation

$${}_3\text{Li}^6 + {}_1\text{H}^2 + \frac{E(\text{D})}{c^2} = {}_2\text{He}^4 + 2 \frac{E(\alpha)}{c^2}, \quad . \quad . \quad (22)$$

or, in numbers,

$$6.0169 + 2.0147 + 0.0001 = 8.0077 + 2 \frac{E(\alpha)}{c^2}.$$

While nuclear transformations by means of fast electrons have hitherto been observed in a few cases only,  $\gamma$ -rays of sufficiently high frequency are more effective. Of particular importance for what

\* For the reaction  ${}_1\text{H}^2(d, p){}_1\text{H}^3$  Döpel obtained the surprising result that the lower limit is experimentally fixed by the least potential required to produce canal rays and not by the nuclear reaction coming to an end. The fact that nuclear reactions do take place, if seldom, at velocities as low as those corresponding roughly to the velocities of gas molecules in the interior of a star at a temperature of a million degrees at last provides us with a source of energy for stellar radiation.



follows is the observation that bombardment of materials containing deuterium gives rise to protons. But what is the residue? If from a particle of mass 2 and charge 1 we remove a proton of mass 1 and charge 1, we are left with an entity having zero charge and mass 1. Here for the first time we meet with a new elementary particle, the neutron. Strong evidence for its existence was already forthcoming from other directions before the above-mentioned observations were made; this evidence we shall discuss in the next section.

### 5. The Neutron.

Up to the year 1932 it was reckoned as among the incontrovertible facts of physics that the positive charge is always associated with the mass of hydrogen, the negative with the mass of the electron, and that these two particles, protons and electrons, form the ultimate constituents of matter. However, in that year, in rapid succession, two new elementary particles were discovered, the neutron mentioned above and the positive electron, or *positron*. Of these, the neutron has revolutionized our ideas of nuclear structure.

In 1930, Bothe and Becker observed that beryllium and other light elements when bombarded by  $\alpha$ -particles, emit not only a hard  $\gamma$ -radiation but also a penetrating radiation with entirely different properties. In substances containing hydrogen this radiation sets free protons of high velocity, although in the Wilson chamber it leaves no traces of itself; that is, the proton tracks, so to speak, start from nothing. If we begin by assuming that here we are again dealing with an extremely hard  $\gamma$ -radiation—owing to the great penetrating power  $\alpha$ - and  $\beta$ -rays are at once ruled out—we are forced to interpret these protons as a Compton effect which has taken place in the nucleus, to which the calculations of p. 688 apply if we replace the mass of the electrons by the mass of the protons. In this way we obtain the value  $55 \times 10^6$  electron volts for the hardness of the  $\gamma$ -rays, while according to the mass relationships the bombardment of beryllium can only yield a much smaller amount of energy. The frequency of these proton tracks, moreover, is about a thousand times as great as we should expect from the assumed Compton effect. Further, a  $\gamma$ -radiation would necessarily set free photo-electrons also; no trace of these is to be found. Finally, the velocity imparted to nitrogen nuclei by the radiation is three times as great as would be produced by a light quantum of  $55 \times 10^6$  electron volts in a Compton effect on the nitrogen nucleus.

We can extricate ourselves from all these difficulties if we assume with Chadwick that this radiation consists of charge-free corpuscles, which owing to the absence of electric charge cannot have any ionizing effect and hence are not retarded by the outer shells of the atom. Such particles do not differ from light quanta, which also have a mass  $h\nu/c^2$ , except that they do not move with the velocity of light.



Their collision with a nucleus is taken as being like the collision of elastic spheres. This means that the forces between the colliding particles decrease according to a very high power of the distance. The nature of these forces is still a matter of speculation. If we calculate the velocities which a proton and a nitrogen nucleus will acquire in colliding centrally (these can both be obtained from their ranges), the principle of the conservation of momentum gives the two equations

$$\left. \begin{aligned} v_p &= \frac{2M_n v_n}{M_p + M_n} \\ v_N &= \frac{2M_n v_n}{M_N + M_n} \end{aligned} \right\} \dots \dots \dots (23)$$

where the indices  $p$ ,  $N$ ,  $n$  refer to proton, nitrogen nucleus, and neutron respectively; hence

$$\frac{v_p}{v_N} = \frac{M_N + M_n}{M_p + M_n} \dots \dots \dots (24)$$

From this equation Chadwick obtained the value  $1.15 \pm 0.1$  for the mass of the neutron.

The *exact* mass of the neutron can be found by comparing masses in various nuclear transformations. The simplest to use is the reaction  ${}_1\text{H}^2(\gamma, n){}_1\text{H}^1$  (see top of p. 752), which has also been observed to take place in the reverse direction,  ${}_1\text{H}^1(n, \gamma){}_1\text{H}^2$ . We have  ${}_1\text{H}^1 = 1.00814$  and  ${}_1\text{H}^2 = 2.01473$ , from observations with the mass spectrograph, taken relative to  ${}_8\text{O}^{16} = 16$ . Also  $h\nu = 2.23$  Mev and  $E_n$  is approximately zero, so that

$${}_0n^1 = 1.00898$$

Owing to the absence of charge on the neutron, which means that it can react only with nuclei, it exhibits very remarkable properties when made to pass through matter. As there are more nuclei in 1 c.c. of paraffin-wax than in 1 c.c. of lead, beams of neutrons are more weakened by paraffin-wax than by an equally thick layer of lead. This serves to distinguish between  $\gamma$ -rays and neutrons. Further, embedding the ionization chamber in paraffin-wax reveals an important effect due to the neutrons, for then beams of protons are shot out by the paraffin-wax and ionize the gas in the chamber in their turn. Neutrons, however, are detected most readily by the fact that they give rise to radioactive isotopes of various elements, in particular, rhodium.

The mass (1.00898) and hence the energy content of the neutron is greater than the sum of the masses of the electron and proton (1.00814), so that the neutron should be unstable with respect to a breakdown into

the other two particles. In fact, such "radioactive decay" of the free neutron has been observed, the half-life being about 12 minutes.

Nuclear disintegrations brought about by beams of neutrons have been observed in great numbers. Curiously enough, the very slow thermal neutrons are often the most effective and very sharp "resonance energies" are found. As might be expected, neutrons are much more effective than charged particles, for the majority of the latter fail to act because their energy is spent in detaching outer electrons from the bombarded atoms.

We may amplify this last point a little. As we saw in the case of the disintegration of lithium, most nuclear processes are highly exothermic reactions. The question naturally arises why these reactions do not spread like the explosion of a mass of gunpowder. The reason is that it is only in an infinitesimal number of cases that the energy of the charged products of disintegration, though very great, can again act effectively in a nuclear collision. The possibility of bringing about atomic transformations on the large scale is therefore exceedingly small. If neutrons are used the conditions are much more favourable, as then the only particles which suffer retardation are those which actually collide with a nucleus. The usual means of producing neutrons by means of a nuclear disintegration yields only a small supply. If, however, there were a process which would set free more neutrons than were shot in to bring it about, the possibility of a chain reaction would exist, provided that the volume of material is so large that only a few neutrons are lost. This possibility was opened up by the discovery in 1939 by Hahn and Strassmann of the fission of the uranium nucleus. Whereas all previous artificial nuclear processes involved the splitting-off of elementary particles only, fission causes the uranium nucleus to break up into two fragments of intermediate weight and a number of neutrons. This will be discussed more fully in § 10.

## 6. The Positron.

Anderson and, independently, Kunze found that when a strong magnetic field was used, cloud chamber photographs of cosmic-ray particles showed tracks of a new type whose curvature corresponded to positive particles, but their intensity to electrons. Confusion with protons is ruled out by the fact that proton tracks so strongly curved could not have the range actually found, and the possibility of a mistake about the direction in which the track was traversed was excluded by making the track pass through a plate of lead, when the side where the curvature is greatest of course represents the later portion of the track. Subsequently a large number of cloud photographs showed that these tracks almost always occur along with an electronic track originating at the same place; in fact, whole "showers" of positive and negative electronic tracks are sometimes found starting from a single point





PLATE III.



Fig. 1.—“ Showers ” of electrons  
and positrons liberated by  
cosmic rays (p. 754)

[*Proc. Roy. Soc. A*, 139, Plate 22]



Fig. 2.—Electron pair, positive  
and negative, released by  $\gamma$ -rays  
(p. 755)

[*Comptes rendus*, 197, p. 237, 1933]

(cf. fig. 1, Plate III). These double tracks may also be produced using  $\gamma$ -rays; however, a photon of over  $10^6$  electron volts is always required (fig. 2, Plate III). Soft  $\gamma$ -rays do not produce the phenomenon.

This minimum value of  $10^6$  electron volts serves as a hint for theory, for it corresponds exactly to the mass of two electrons. It is therefore feasible to interpret the occurrence of positive and negative electrons as meaning that a light quantum becomes materialized, transforming itself into one or more "electron pairs". The conservation of charge and energy is guaranteed, but the conservation of momentum cannot hold, the reason being that the light quantum of all particles of the same energy (rest energy + kinetic energy) is that with the greatest momentum. For by formula (72) on p. 254 and formula (82) on p. 257, the momentum of a material particle of rest-mass  $m_0$  is

$$p_m = \frac{m_0 v}{\sqrt{(1 - v^2/c^2)}}, \quad \dots \dots \dots (25)$$

and its energy is

$$E_m = \frac{m_0 c^2}{\sqrt{(1 - v^2/c^2)}}, \quad \dots \dots \dots (26)$$

so that

$$\left(\frac{p}{E}\right)_m = \frac{v}{c^2} = \frac{v}{c} \cdot \frac{1}{c}; \quad \dots \dots \dots (27)$$

for the light quantum, on the other hand,

$$p_l = \frac{h\nu}{c}, \quad E_l = h\nu,$$

so that

$$\left(\frac{p}{E}\right)_l = \frac{1}{c} > \frac{v}{c} \cdot \frac{1}{c} \quad \dots \dots \dots (28)$$

Hence the phenomenon can occur only in the presence of a nucleus which will take up the surplus momentum. The small energy transfer connected therewith may be neglected.

Apart from this "materialization" of light quanta, however, Curie and Joliot have observed that positrons also appear when aluminium is disintegrated by  $\alpha$ -particles. In this process protons of very great range had previously been obtained as a product. Now a process in which a neutron and, subsequently, a positron are emitted clearly exists as well. Later, positrons were recognized to be entirely normal accompaniments to nuclear changes.

Now why is the positron found less frequently than the electron? Even before the discovery of the positron Dirac had developed a generalized wave-mechanical theory of the electron (p. 711) in which the spin appears quite naturally. We could not go into this theory in

detail, owing to the advanced nature of the mathematics involved. Formerly it was regarded as a defect in the theory that it permitted the appearance, even if only for the time being, of positive electrons; but these very entities have now been observed. The annihilation of the positrons obviously results from their union with electrons with the emission of  $\gamma$ -radiation. Here the momentum theorem can only be satisfied if two or more quanta are produced simultaneously. Observations on the passage of  $\gamma$ -rays through matter make it probable that materialization and dematerialization of the  $\gamma$ -quantum play a part in addition to normal absorption.

## 7. Artificial Radioactivity.

Curie and Joliot made the discovery that the emission of positrons from aluminium continues for some little time after the bombarding radiation is cut off, dying away exponentially. They interpreted this as meaning that when aluminium is bombarded with  $\alpha$ -rays an unstable nucleus is formed which is transformed into a stable one according to the laws of atomic disintegration.

Subsequently this *induced radioactivity* was found to occur in hundreds of nuclear transformations. This is not surprising, for in most cases the balancing of the masses gives isotopes which differ from those already known. They are not to be found in nature because, if they ever did exist, they must long since have been transformed into the stable forms.

We can, in fact, understand the nuclear transformations discussed in § 4 (p. 747) better if we assume that in every case the bombarding particle is captured to form an unstable intermediate nucleus, which in these cases has a life too short to measure.

Atoms rendered artificially radioactive have found many important applications in medicine and in industry. For example, by administering small amounts of radioactive sodium to a patient, its progress through the blood stream can be followed by means of a Geiger counter. Similar experiments can be performed with plants in order to find out how they assimilate and distribute foods. In engineering, the rate of wear of bearings has been determined by previously introducing a small amount of a radioactive metal into the bearing metal and then measuring the activity of the lubricating oil.

In all artificially radioactive elements the emission of either electrons or positrons is observed and sometimes  $\gamma$ -radiation as well. The reason for the preponderance of the first two phenomena, collectively referred to as  $\beta$ -transformations, will be made clear in what follows. The occurrence of  $\gamma$ -rays accompanies many processes. In addition, there is a process in which the nucleus captures an external electron. One might expect that a positron ejected from the nucleus in a  $\beta^+$ -process could combine with one of the external electrons. If this were



the case, one would always observe  $\gamma$ -radiation of 0.5 Mev energy as described above for positron annihilation. However, this radiation is never found, and only the K series of the element in question is observed. It must then be assumed that in this process a K electron plunges into the nucleus and then, as is usual in K emission, the place thus vacated is filled by another electron from the L, M, or some higher shell. However, in order to preserve the energy balance in such *K capture*, it has become necessary to assume the existence of another kind of particle, the *neutrino*, which has only recently become available to direct observation. This will be discussed further in § 13.

### 8. The Role of Protons and Neutrons in Nuclear Structure.

The neutron opens up a new way of explaining the difference between atomic weight and nuclear charge, and thus enables us to drop the assumption of nuclear electrons. The following are the chief objections to nuclear electrons:

(1) The small magnetic moment of the nucleus, already mentioned on p. 744, which according to all the observations is of the order of a nuclear magneton, not of an electronic magneton.

(2) From other observations, the proton and probably every elementary particle has a spin of  $1/2$ . Experimentally it is known (p. 744) that nuclei of even atomic weight have spins of even number, nuclei of odd atomic weight spins of odd number. This is incompatible with the assumption of nuclear electrons. For if  $Z_p$  is the number of protons, the assumption that the nucleus is built up of protons and electrons means that the atomic mass  $A$  is equal to  $Z_p$  and that  $Z_e$ , the number of electrons in the nucleus, is equal to  $Z_p - Z$ . The sum of protons and electrons is  $Z_{p+e} = 2Z_p - Z = 2A - Z$ . Whether this is even or odd, then, depends on the atomic number  $Z$ , not on the atomic mass  $A$ . On the other hand, by the laws of combination of spins given on p. 669, an even number of particles always gives a total spin which is a whole number, and an odd number of particles a total spin which is a multiple of  $1/2$ . Hence  ${}_1\text{H}^2$  and  ${}_7\text{N}^{14}$ , which have odd atomic numbers, would each have a spin which is a multiple of  $1/2$ , whereas it is actually found to be 1 in both cases.

(3) According to the indeterminacy relation, if the electron is to be restricted to a space of nuclear dimensions it must have a momentum of the order of  $6 \times 10^{-15}$  gm. cm./sec. and hence an energy of about 100 Mev. Thus, although we obtain for the electronic shells term values of the energy states of the correct order of magnitude, the values in the case of the nucleus are too high by more than a factor of 10.

But how are we to explain  $\beta$ -disintegration if there are no electrons in the nucleus? The following account of the state of affairs may be plausible by analogy. If an atom emits a quantum of visible light we

do not conclude from this that the electronic shells contain light quanta, but on the contrary that the light quantum originates during the passage of the electronic shells to a lower energy state. Now the energy relationships in the nucleus are such that the energy set free in the passage to a lower state is sufficient to cover the rest mass of the electrons and their velocities. We therefore assume that *the emission of a nuclear electron is associated with the transformation of a neutron into a proton, and the emission of a positron with the transformation of a proton into a neutron*. But, as in the case of the emission of light, we cannot attempt to describe exactly how the process takes place. The charged particle emitted then serves to balance the charges. Instead of the emission of positrons, however, an electron of the K shell may be captured by the nucleus, in which case only lines of the K series are emitted (p. 654).

If we assume that nuclei are built up of protons and neutrons we obtain the energy of binding as the equivalent of the total mass of all the protons and neutrons less the actual nuclear mass.\* This is always less than the sum of the masses of the individual neutrons and protons; that is, when they combine to form a nucleus, energy is set free. Otherwise, of course, nuclei would not be stable. The matter may be illustrated in a particularly clear way by means of a graph: Relative to a set of axes in three dimensions we plot  $Z_p$ , the number of protons (which is equal to the charge on the nucleus) along the  $x$  direction;  $Z_n$ , the number of neutrons (which is equal to the difference between the atomic mass and the atomic number) along the  $y$  direction; and the (negative) energy of binding downward in the  $z$  direction, so that the nuclei of least energy lie lowest. In this way we obtain a surface, although it is true that we know only a narrow strip of it corresponding to the known natural and artificially produced isotopes. But from this we can surmise that the surface has the form of a trough falling away from the origin. The bottom of this valley contains the stable nuclei and at first runs diagonally in the first quadrant of the  $xy$  plane, the origin being the highest point (fig. 1).

If we look at the numerical values of the energies of binding we see that they run from 0 to 1800 Mev. Since  $932 \text{ Mev} = 1 \text{ a.m.u.}$ , this corresponds to a change in atomic weight of up to 1.9 units. The fact that pure isotopes have atomic weights which are approximately whole numbers when all the values intermediate between 0 and 1.9 can occur is due to the fortunate choice of  $O = 16$  as a standard of atomic weight. As the atomic weights of the proton and the neutron exceed unity by 0.008 or 0.009, 230 particles will make the difference mount up to 1.9, so that the mean binding energy per particle is about 0.0085. Hence,

\* In computing this difference, the mass of the H atom is to be used in place of the proton mass, and in any case the mass of the entire atom rather than merely the nuclear mass. (See p. 763.)



as the surface falls away fairly evenly (see fig. 2), the atomic weights are approximately whole numbers.

The dots in fig. 1 represent the known stable isotopes. At the start, the dots lie on the diagonal line in the first quadrant because the atomic weight is twice the atomic number for the light nuclei, i.e. the number of neutrons is equal to the number of protons. All points lying in the

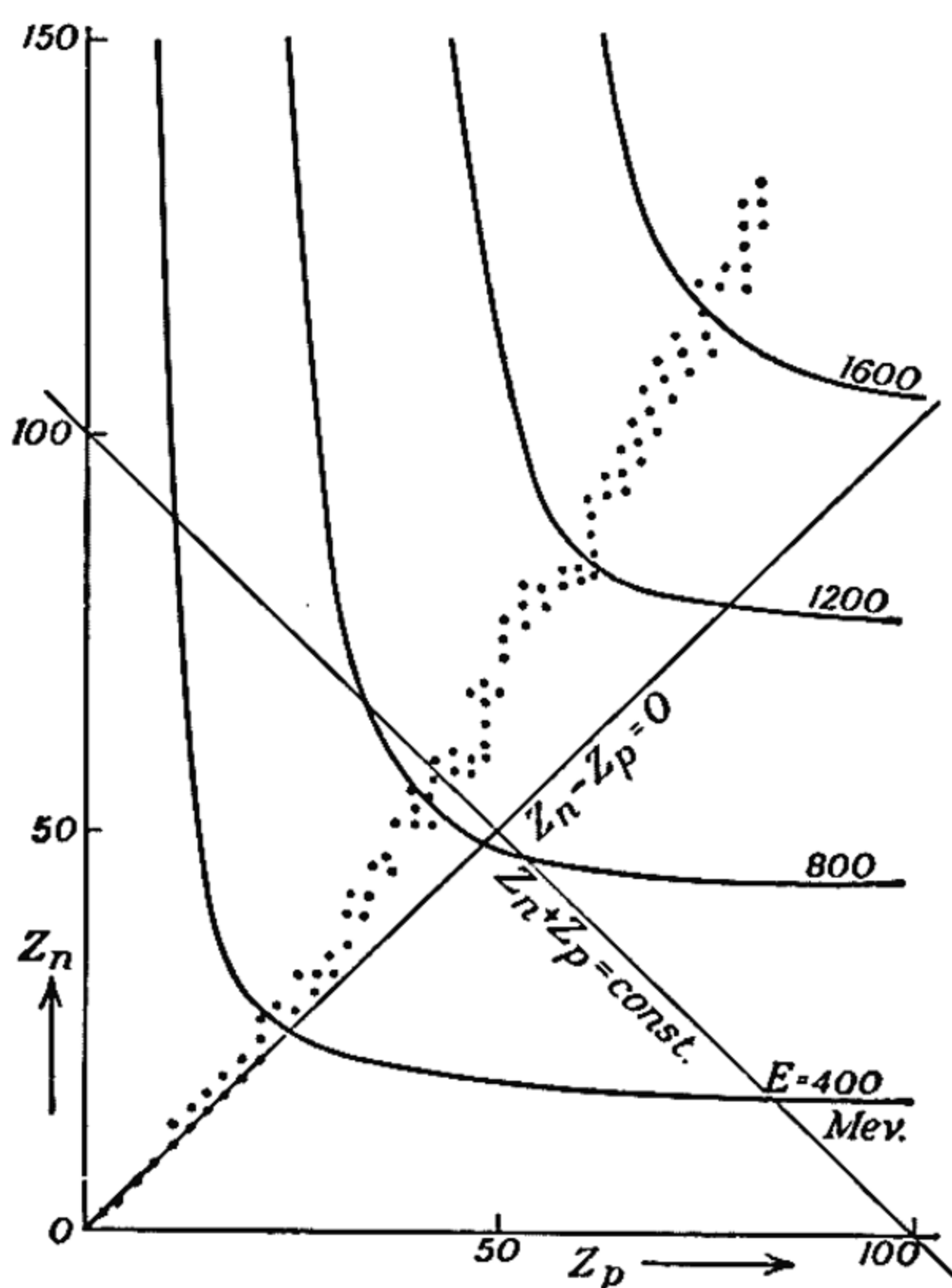


Fig. 1

plane  $Z_p = \text{const.}$  parallel to the  $Z_n$  axis and perpendicular to the plane of the paper have the same nuclear charge—they are isotopes. All points lying on a vertical plane ( $Z_n + Z_p = \text{const.}$ ) parallel to the diagonal of the second quadrant of the  $xy$  plane have the same mass—they are isobars.

This method of approach enables us to deduce much information about the stability of nuclei relative to the transformations we know to occur on the emission of  $\alpha$ -rays, electrons or positrons (both of the latter referred to as  $\beta$ -transformations). Only the state of least energy corresponds to a system in equilibrium. In thermodynamics it was emphasized that it is the minimum of the *free* energy, not of the total energy, that is involved; nevertheless, this distinction is not important in nuclear reactions taking place on earth because the thermal energy



is negligible compared with the nuclear energy, which means that we are practically at absolute zero, so far as nuclear reactions are concerned. Thermodynamic considerations, however, are quite relevant in the interior of stars, where temperatures of millions of degrees are believed to exist. In  $\alpha$ -disintegration, where both  $Z_p$  and  $Z_n$  are diminished by 2, the magnitude of the energy  $E$  must satisfy the relation

$$E(Z_p, Z_n) > E(Z_p - 2, Z_n - 2) + E(2, 2). \quad (29)$$

As we may regard the change by 2 units approximately as a differential, we may write instead

$$2 \left( \frac{\partial E}{\partial Z_p} + \frac{\partial E}{\partial Z_n} \right) > E(2, 2), \quad (30)$$

i.e. the steeper the fall of the energy surface, the more assured is the stability. If, for the known nuclei,  $E$  is now expressed as a function of the atomic weight, this corresponds to the intersections of the bottom of the valley with the planes  $Z_p + Z_n = \text{const.}$  We thus obtain fig. 2. From it we see that, beginning with atomic weights of about 120, the inclination of the surface decreases, and  $\alpha$ -disintegration occurs only among heavy nuclei.

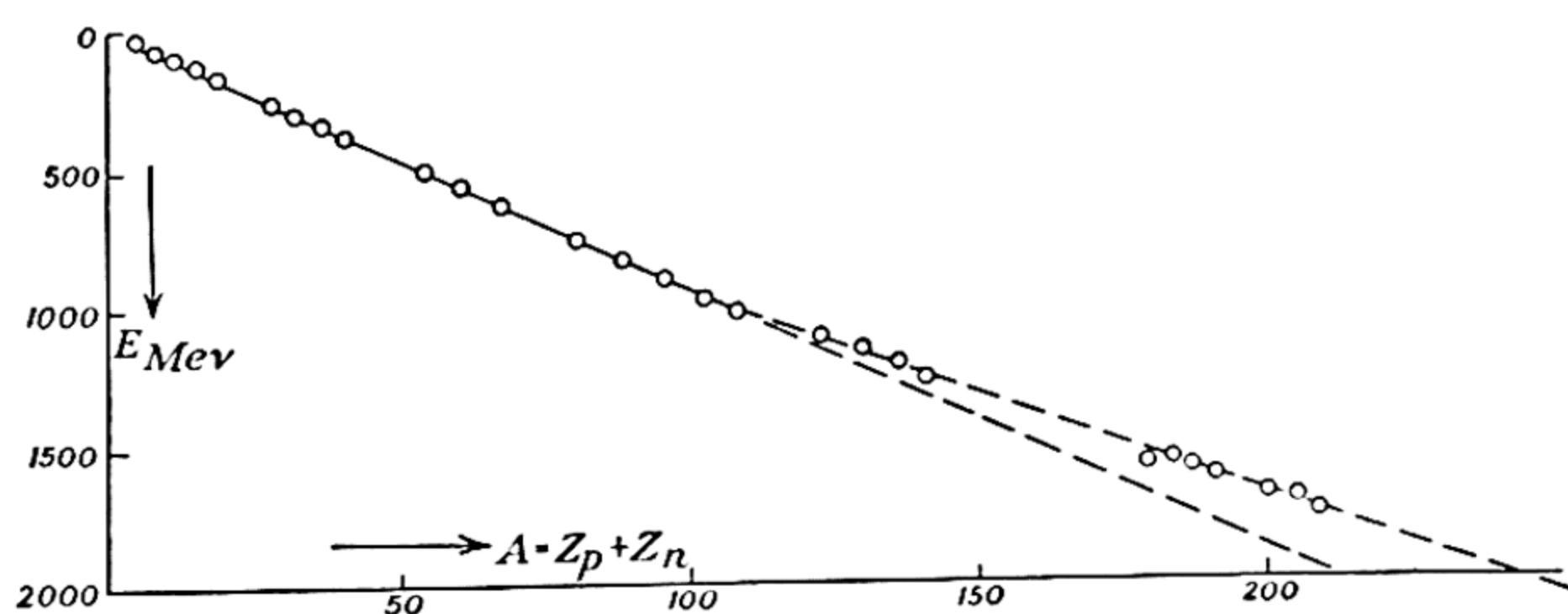


Fig. 2

In order to discuss the stability of  $\beta$ -disintegrations, the surface concept must be further refined. The energies of binding calculated from the atomic weights are such that with a more accurate plot we can no longer make all the points lie on a smooth surface; the surface becomes crinkled. If, however, we join the points for which  $Z_p$  is even and  $Z_n$  is even ( $A$  even), those points for which  $Z_p$  is odd and  $Z_n$  is odd ( $A$  even), and, finally, the points for which  $A$  is odd (either  $Z_p$  or  $Z_n$  odd), we obtain three smooth surfaces which we denote \* by  $E_{ee}$ ,  $E_{ue}$  and  $E_{uu}$ . Then an isobaric section gives a single curve of inter-

\* It is not necessary to distinguish between  $E_{eu}$  and  $E_{ue}$

section, or two curves, according as it corresponds to an odd or even atomic weight. This is shown for an even atomic weight in fig. 3a. The  $E_{uu}$  curve lies above the  $E_{ee}$  curve, from which we conclude that in general nuclei for which  $Z_p$  and  $Z_n$  are odd are unstable with respect to  $\beta$ -transformations, since, of course, the next lower point can be reached by such a change. Actually there are only four nuclei of this type:  ${}^1_1\text{D}^2$ ,  ${}^3_3\text{Li}^6$ ,  ${}^5_5\text{B}^{10}$ ,  ${}^7_7\text{N}^{14}$ . The existence of these nuclei is possible because at the start the walls of the nuclear energy surface are not far apart and are steep, so that we have the stage of affairs shown in fig. 3b. The favouring of the type  $ee$  is rightly associated with pairwise cancelling of the spins. As the  $\beta$ -disintegration takes place in an isobaric section, it is only a question of whether there is a nucleus lower

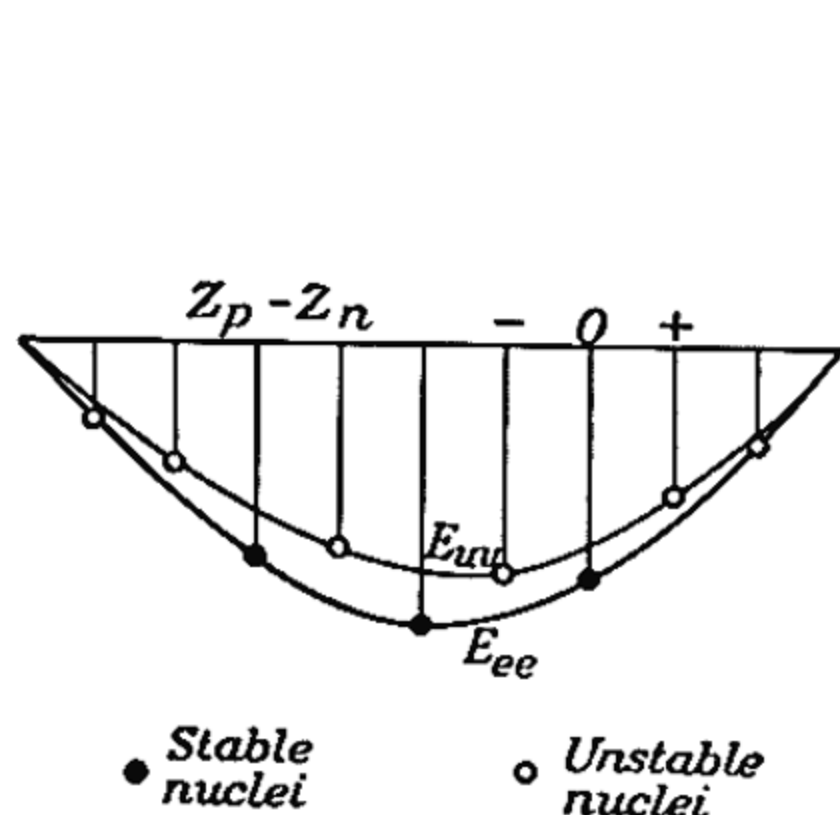


Fig. 3a

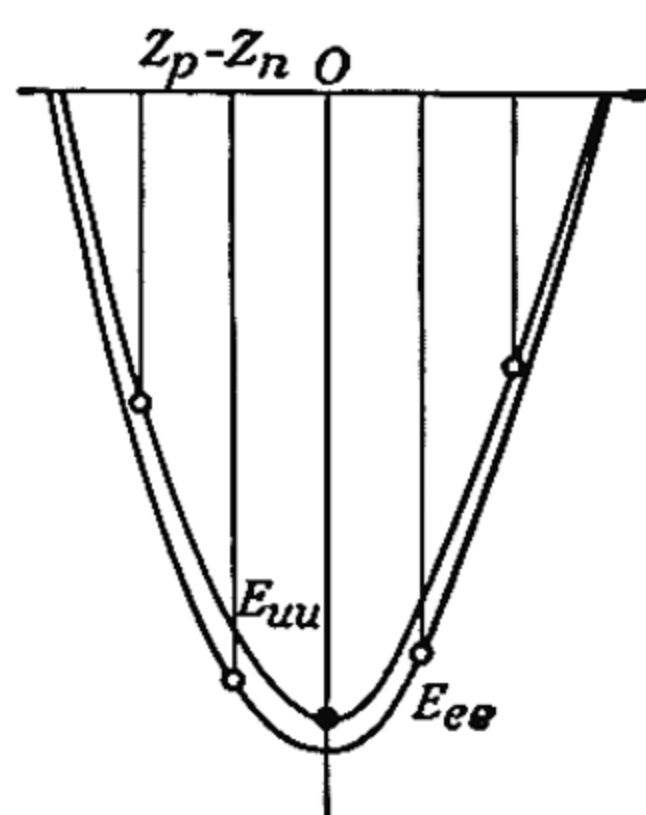


Fig. 3b

down in the immediate neighbourhood—at any rate, double transformations with emission of two electrons have not so far been observed. If we make the section through a nucleus of the type  $ue$ , there are other nuclei at unit distance, which are  $\beta$ -unstable since they lie higher. Hence *if the atomic weight is odd, there are no stable isobaric substances*. This is Mattauch's first law of isobars.

Since the  $E_{uu}$  surface has no stable nuclei beyond  ${}^7_7\text{N}^{14}$ , all the points corresponding to stable nuclei of even atomic weight must lie on the  $E_{ee}$  surface, but they are 2 units apart. That is, a  $\beta$ -disintegration results in passage to the  $E_{uu}$  surface, with the formation of a new unstable nucleus. From the position of the two surfaces we see from fig. 3a that here 3 stable nuclei are possible, while the two next ones, with the relationships we have assumed, have a lower neighbour in the  $E_{uu}$  surface, i.e. they are  $\beta$ -unstable. We thus obtain Mattauch's second law of isobars: *If the atomic weight is even there are several stable isobaric substances, but they are never those for which the nuclear charge differs by unity.*

In order to arrive at a statement about the possible isotopes, we

project the points of the energy surface and the bottom of the valley on to the  $xy$  plane. We can immediately omit all the nuclei of type  $uu$  as unstable (in fig. 4, in which the stable nuclei which actually occur are shown by black circles, the unstable nuclei are crossed out). If, for example, we consider the isotopes of Br, the two nearest the bottom of the valley are  ${}_{35}\text{Br}^{79}$  and  ${}_{35}\text{Br}^{81}$ . If we describe isobaric sections through them, within which  $\beta$ -transformations take place, we see that they lie lowest in these also, as they are nearest to the bottom of the valley. The nearest possible isotopes, however, lie higher in their

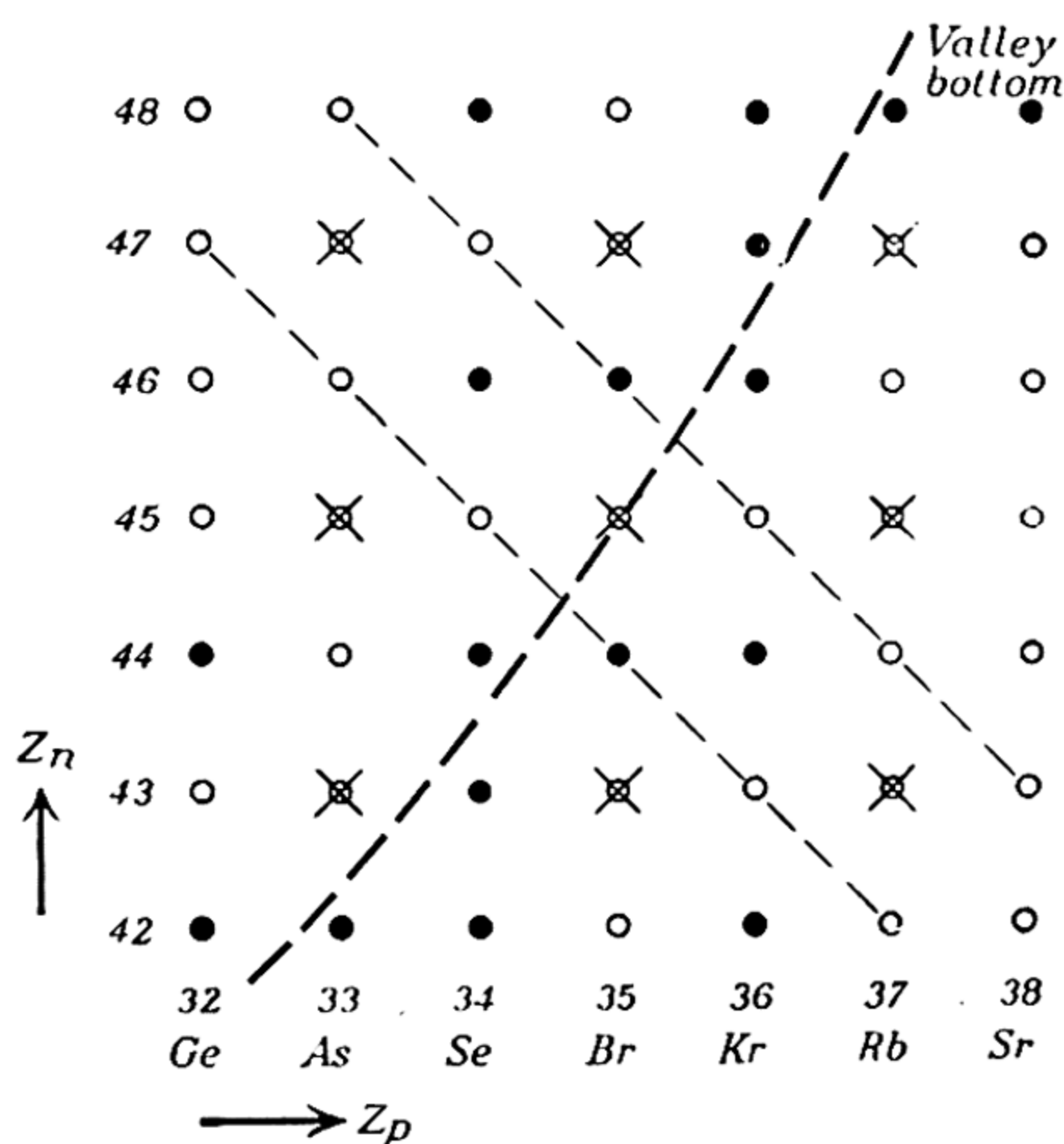


Fig. 4

isobaric sections than other nuclei, i.e. they are  $\beta$ -unstable. We thus obtain Aston's isotope rule: *elements of odd atomic number (such as Br) have at most two stable isotopes*. In the same way, we see immediately from the figure that a larger number of stable isotopes is possible when the nuclear charge is even, especially when the  $E_{ee}$  and  $E_{ue}$  surfaces are far apart. A good many of the  $\beta$ -unstable atoms marked by empty or crossed-out circles actually have been found in nuclear transformations, so that now we know something of the region around the valley bottom.\*

By means of the energy surface we can also explain the remarkable "resonance energies" mentioned in §5 as being observed in the binding of neutrons, i.e. particularly effective velocities which often

\* See, for example, Sullivan, W. H., *Trilinear Chart of Nuclear Species*, New York, Wiley, 1949.



amount to only a few electron volts and thus are of quite a different order of magnitude from the nuclear energies. If we go along the valley of stable nuclei from one isobaric section to the next, the surface falls by about 8 million electron volts. In adding a neutron we are, it is true, not moving in this direction of greatest fall, but in the isotopic section; however, the new nucleus will in any case always lie several million electron volts lower. On the other hand, an excited state of this nucleus may come to about the same level as the original nucleus, and this energy of excitation corresponds to the order of magnitude of the nuclear energies within a few million electron volts. If now the approaching neutron has sufficient kinetic energy to supply exactly the amount needed to make the energies correspond, binding takes place without any accompanying phenomena. The excited nucleus then passes to its normal state with the emission of  $\gamma$ -rays and there follows a  $\beta$ -transformation to the lowest point of the isobaric section in which the new nucleus lies. Excited nuclei, like excited atoms, can be metastable, and then have a life which is long in comparison with that of metastable atoms, amounting even to hours. This explains the fact of nuclear isomerism—the occurrence of nuclei of the same atomic weight and the same charge, but with differing lengths of life.

The representation by means of the nuclear energy surface, although very easy to follow, does not give a perfectly accurate description of the stability relationships in the isobaric section. It is on the difference in atomic weight that the relative stability of two nuclei depends (see Ex. 135). Owing to the slight difference between the mass of the hydrogen atom and that of the neutron, however, the difference in energy between two nuclei in the same isobaric section is not exactly equal to the difference in atomic weight, for in advancing along the isobaric section the ratio of the number of protons to the number of neutrons changes. Hence in critical cases it is advisable to plot the deviation of the atomic weight from integral values over the  $ZN$  plane in the various isobaric sections. If this is done, however, the various isobaric sections can no longer be brought into relation with one another.

*Ex. 135.* Show that a nucleus of atomic mass  $A$  and nuclear charge  $Z$  is stable as regards  $\beta$ -transformation and transformations involving  $K$ -capture if its atomic weight is less than that of its neighbours  $Z + 1$  and  $Z - 1$ . (Write down the equation for the conservation of energy, taking into consideration the mass of any electrons produced or destroyed.)

## 9. The Liquid Drop Model of the Nucleus.

If we really want to calculate the binding energies and the stationary states of the nuclei, we must make a model of the configuration of the nuclear components. Here, unlike the situation for the atom as a whole, there is no central body, so that the planetary model is not appropriate.

The similarity between the constituents — the difference between charged and uncharged particles does not play a big part here and is neglected in what follows—suggests that we take a liquid drop as our model. To explain the very strong cohesion of this droplet we have to introduce a new kind of “nuclear force” acting between the constituents, no matter whether they are charged or not. The fact that the energy always comes to about the same amount, about 8 Mev for each new particle taken in, shows that the forces (like those binding a liquid together) fall off rapidly with the distance, so that only the interactions with immediately neighbouring particles are important. With forces which, like electrostatic forces, diminish relatively slowly with distance, the energy of the mutual action between the newly arrived particle and all those previously there must be taken into account, so that the binding energy per particle must rise rapidly with the number of particles already present (see below). In actual fact, experiments on the scattering of fast protons by protons have shown that for very close approaches there is a deviation from the classical scattering formula derived from the laws of electrostatics, which proves that, contrary to the Coulomb repulsion, there exists a force of *attraction* which decreases rapidly with distance.

The droplet model even yields the finer details of the situation, for example the existing slight progressive change in the binding energy per particle, i.e. the sum of all the particle masses minus the nuclear mass, all divided by the nuclear mass, or the quantity  $\delta M/A$ . If we think of the constituents of the nucleus as in the closest spherical packing, each particle has 12 neighbours, and the energy set free by bringing in another particle is  $6U_0$ , where  $U_0$  is the binding energy between two particles. From this, however, we have to subtract two quantities: (1) a diminution of the binding energy of the particles lying on the surface, each of which interacts with only six neighbours, and (2) the electrostatic energy of the protons which owing to mutual *repulsion* has the opposite sign to the energy of nuclear *attraction*. So far as the former is concerned, we can consider its variation with the number of particles  $A$ , as follows: The volume increases in proportion to  $A$ , so that the radius of the droplet increases as  $A^{1/3}$ . Its surface, therefore, varies as  $A^{2/3}$ , and the contribution of this quantity per particle is proportional to  $A^{2/3}/A = A^{-1/3}$ .

For the electrostatic energy, the slow decay of the forces makes it necessary to calculate the energy of a sphere of radius  $a$  having a uniform charge density  $\rho$  as in Ex. 80, p. 290. The energy is proportional to  $\rho^2 a^5 = \rho^2 A^{5/3}$ . Since the ratio of protons and neutrons does not change much with atomic weight, we may regard  $\rho$  as constant to the degree of approximation sufficing here, and the contribution per particle becomes approximately proportional to  $A^{2/3}$ . Hence the relation between  $A$  and the binding energy per particle is expressed by



a formula with three constants,

$$-\frac{\delta M}{A} = 6U_0 - K_0 A^{-1/3} - K_c A^{2/3}. \quad (31)$$

With the empirically determined values of  $U_0$ ,  $K_0$ , and  $K_c$  this formula expresses the relation surprisingly well, as shown by fig. 5 (p. 766).

The left-hand scale gives the binding energy per particle expressed in millions of electron volts and in thousandths of an atomic mass unit. The right-hand scale is the so-called *packing fraction*, given in ten-thousandths of a unit. This quantity varies in close correspondence with the binding energy per particle.\*

The increase in the electrostatic energy has the further consequence that, with increasing atomic weight, the lowest energy state of a nucleus no longer occurs when the number of neutrons equals the number of protons. Instead, for the stable nuclei, the number of neutrons increases more and more with respect to the number of protons until finally only  $\alpha$ -unstable nuclei are possible.

We shall refine equation (31) somewhat and at the same time put it into a form which makes it easier to recognize the meaning of the constants. In order to account for the slopes of the sides of the nuclear energy surface we must add a term representing the neutron-proton ratio. The lighter stable nuclei have  $N = Z$ , while for the heavier ones there is a progressive increase in the neutron excess. From this we conclude that the binding energy is greatest for  $N = Z$ , and that the deviations in the total energy are attributable to the increase in the electrostatic energy. Consequently, we add a term in  $(N - Z)^2$ . By analogy with the case of binary solutions in physical chemistry, this energy term will depend only on the *relative* concentrations, so that it will take the form  $b(N - Z)^2/(N + Z)^2$ . Further, we represent the nuclear radius by

$$r = r_0 A^{1/3}. \quad (32)$$

Finally, we insert a term representing the splitting of the energy surfaces attributable to the pairing of the spins. This term is somewhat harder to justify theoretically. Using  $u = 6U_0$ , the result is

$$\frac{\delta M}{A} = -u_0 + b \left( \frac{N - Z}{N + Z} \right)^2 + c A^{-1/3} + \frac{3Z^2 e^2}{5r_0} A^{-4/3} + \gamma k A^{-1/4}, \quad (33)$$

\* It may be useful to summarize the relevant definitions at this point. One *a.m.u.* is  $\frac{1}{18}$  of the mass  $O^{16}$ , and the *isotopic mass* of an atom species is its atomic mass expressed in *a.m.u.* The *mass number* is the nearest whole number to the isotopic mass, and all known isotopic masses fall within 0.1 of the closest integer. The small difference is the *mass defect* (or *mass correction*). Finally, the *packing fraction* is the mass correction divided by the mass number.



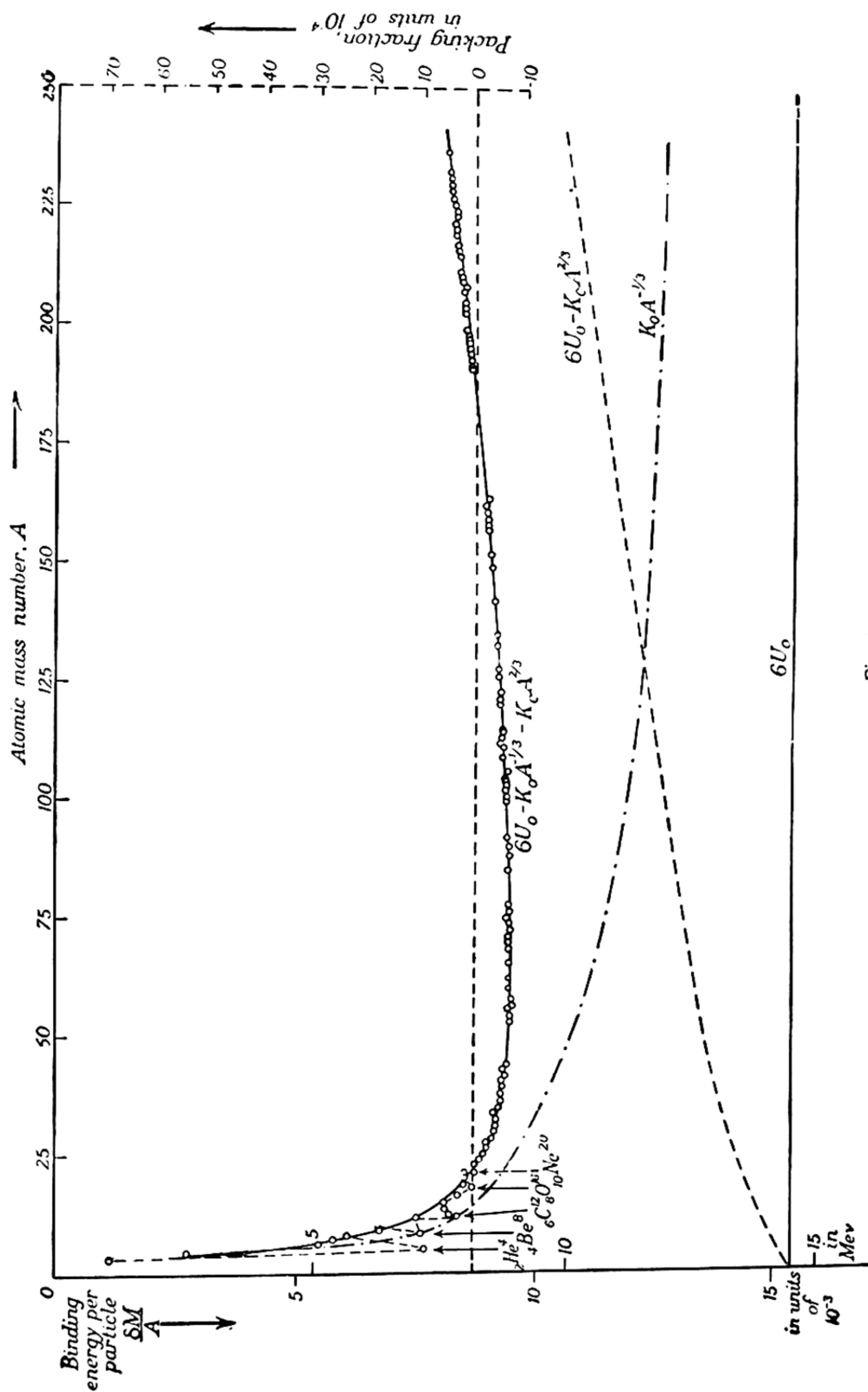


Fig. 5

where

$$\begin{aligned}\gamma &= +1 \text{ for } uu, \\ &= 0 \text{ for } ue, \\ &= -1 \text{ for } ee.\end{aligned}$$

The numerical values of the constants obtained after smoothing out the small irregularities of the nuclear energy surface are, expressing  $\delta M$  in thousandths of a mass unit,

$$\left. \begin{aligned}u_0 &= 15.7, \quad b = 22, \quad c = 16.5, \\ \frac{3e^2}{5r_0} &= 0.646, \quad k = 36.\end{aligned} \right\} \quad (34)$$

The constant in the nuclear radius formula (32) has the value

$$r_0 = 1.41 \times 10^{-13} \text{ cm.} \quad (35)$$

curiously, this is the radius of a sphere over which the charge  $e$  could be spread uniformly to have the electromagnetic energy corresponding to the mass of the electron. The "classical electron radius" is represented by  $2r_0$ .

It must be mentioned that the energy  $u_0$  representing the binding energy of a nucleon involves not only potential energy [represented by the negative term in equation (33)] but a kinetic part (positive terms) as well.

## 10. The Fission of Uranium.

This process, mentioned at the end of § 5, may be visualized remarkably well by using the liquid drop model of the nucleus. Firstly, because of the existence of surface energy, the most favourable form is the sphere. However, because of Coulomb forces, a more elongated shape may become more advantageous inasmuch as this allows the charges to be farther apart, on the average. In fact, a splitting into two spheres can result in an increase in energy. As for water drops, supplying energy can bring about vibrations that result in a necking-in of the drop near the middle. The question arises as to when such a situation will become unstable, leading to actual splitting into two fragments.

In order to investigate the stability we develop the functions of the radius—surface area, volume, electrostatic energy—according to powers of a function of the polar angle  $\theta$ . With departures from the spherical form,  $r$  will depend on  $\theta$ , but we still assume that the drop has symmetry of rotation. Instead of the angle  $\theta$  itself, we use spherical harmonics—simple polynomials in  $\cos \theta$  (see pp. 819–824). The advantage is that these functions represent solutions of the equation of vibration of the drop. Because of the symmetry of the vibrating drop,

$f(\pi - \theta)$  must equal  $f(\theta)$ , and the uneven powers of  $\cos \theta$  must be absent. Hence we may write

$$r' = r(1 + a_0 P_0 + a_2 P_2(\theta) + \dots), \quad . \quad . \quad . \quad (36)$$

where  $P_0 = 1$  and  $P_2 = \frac{1}{2}(3 \cos^2 \theta - 1)$ . Multiplying by  $2\pi \sin \theta d\theta$  and integrating from 0 to  $\pi$  gives us the value of the surface of the deformed drop. The result, carried to the second order, is

$$F' = 4\pi r^2(1 + 2a_0 + \frac{4}{5}a_2^2). \quad . \quad . \quad . \quad (37)$$

Then the surface energy amounts to

$$E_0 = cA^{2/3}(1 + 2a_0 + \frac{4}{5}a_2^2). \quad . \quad . \quad . \quad (38)$$

In the same way, the volume of the drop comes out at

$$V = \frac{4\pi}{3}r^3(1 + 5a_0 + \frac{3}{5}a_2^2). \quad . \quad . \quad . \quad (39)$$

Inasmuch as the nuclear "fluid" must be considered incompressible,

$$a_0 = -\frac{1}{5}a_2^2. \quad . \quad . \quad . \quad . \quad (40)$$

Then, according to equation (38) the surface energy becomes

$$E_0 = cA^{2/3}(1 + \frac{2}{5}a_2^2). \quad . \quad . \quad . \quad . \quad (41)$$

In the expression for the electrostatic energy, the mean of the reciprocal of the distance between charges,  $\overline{1/r}$  is to be replaced by  $\overline{1/r'}$ . After some computation, this is found to be

$$\overline{1/r'} = \overline{1/r} (1 - \frac{1}{5}A_2^2). \quad . \quad . \quad . \quad . \quad (42)$$

We now have for the total change in energy

$$\Delta E = \Delta E_0 + \Delta E_{st} = \frac{2c}{5}A^{2/3}a_2^2 - \frac{3}{25} \frac{Z^2 e^2}{r_0 A^{1/3}} a_2^2. \quad . \quad (43)$$

If the second term exceeds the first in magnitude, the slightest disturbance will succeed in splitting the nucleus. The critical condition, determined by equality of the two terms, is given by

$$\left(\frac{Z^2}{A}\right)_{\text{crit}} = \frac{2c}{3e^2/5r_0}; \quad . \quad . \quad . \quad . \quad (44)$$

and, using the numerical values from (34) and (35), the resulting value of the above fraction is about 50. Since  $Z$  is equal to about  $2A/5$  for the heaviest nuclei, the values of the critical atomic mass and critical atomic number turn out to be

$$A_{\text{max}} \approx 312 \quad \text{and} \quad Z_{\text{max}} \approx 125.$$



It appears, then, that actual nuclei having a  $Z < Z_{\max}$  will require excitation in order that oscillation and consequent necking-in take place. The amplitude of this oscillation will determine whether the energy state so attained is stable or not. It seems reasonable to express this energy of excitation as the product of the surface energy by a function of the ratio  $(Z^2/A)/(Z^2/A)_{\text{crit.}}$ . The function may be specified for the simple special cases of (a) weak charge and (b) proximity to the critical charge; and by means of skilful interpolation Alvarez and Bohr were able to trace the entire course of the function to some degree. The results show that the nucleus  ${}_{92}\text{U}^{236}$  requires an excitation energy of 6.8 Mev, while  ${}_{32}\text{U}^{239}$  requires 7.1 Mev. According to the formula for binding energy, the formation of  $\text{U}^{236}$  by the addition of a neutron to  $\text{U}^{235}$  liberates 6.81 Mev; on the other hand, formation of  $\text{U}^{239}$  by the combination of a neutron with  $\text{U}^{238}$  releases only 5.37 Mev. The considerable difference results from the fact that  $\text{U}^{236}$  is of *ee* type, while  $\text{U}^{239}$  is *ue*. Thus the critical factor here is the  $\gamma$ -term in the nuclear-energy formula, which determines the splitting of the energy surface. It turns out that  $\text{U}^{235}$  can be split even by thermal neutrons, whose energy is practically zero, while splitting of  $\text{U}^{238}$  requires fast neutrons with an energy of about 1.5 Mev. Since the fission products have too many neutrons compared with their nuclear charge, free neutrons appear in the process and these are able to propagate a chain reaction.

### 11. Liquid Drop Treated as a Fermi Gas.

At the end of § 9 it was pointed out that  $u_0$  is the difference between the potential energy and the kinetic energy which is also present in the nuclear fluid. As with the electron shells, we get some idea of the distribution of the kinetic energy by treating the nucleus simply as a Fermi gas. It is true that the masses are considerably greater, so that according to p. 632 it might be doubted that the conditions for a Fermi gas obtain here. On the other hand, the density is so much greater that this circumstance is more than compensated. Each nucleon, and indeed every true elementary particle, must be assumed to have spin 1/2, which can appear in either of two orientations. If we differentiate between protons and neutrons, each cell may be occupied by two particles. For a first approximation the difference between proton and neutron may be overlooked and each cell of phase space may be occupied by four particles.\*

We make a very simple assumption about the shape of the potential energy curve, corresponding to the short range of the nuclear forces: Let a nucleon have zero potential energy when beyond the nuclear

\* In a more refined calculation, the existence of electrostatic forces requires that protons and neutrons be assigned to separate cells, each cell being occupied by only two particles.



The numerical value is but slightly different from the value of  $u_0$ , but with positive sign. Here, then, as in the case of the hydrogen atom, the kinetic energy turns out to be about half the potential, but in this instance the relation is more a matter of coincidence.

In the more refined computation which takes into account the differing values of potential for charged and uncharged particles, the number of neutrons comes out to be

$$N = \frac{8\pi v}{3h^3} p_{n, \max}^3 \quad . \quad . \quad . \quad . \quad . \quad (49)$$

and the number of protons

$$Z = \frac{8\pi v}{3h^3} p_{p, \max}^3 \quad . \quad . \quad . \quad . \quad . \quad (50)$$

The sum must be  $A$ . From (47), (49) and (50),

$$p_{n, \max} = \sqrt[3]{\frac{2N}{A}} p_{\max}; \quad p_{p, \max} = \sqrt[3]{\frac{2Z}{A}} p_{\max}; \quad . \quad . \quad (51)$$

and the total kinetic energy becomes

$$\begin{aligned} E_{\text{kin}} &= \frac{3}{10M} (N p_{n, \max}^2 + Z p_{p, \max}^2) \\ &= \frac{3}{20M} \left( \frac{h}{2\pi r_0} \right)^2 \left( \frac{9\pi}{8} \right)^{2/3} A \left[ \left( \frac{2N}{A} \right)^{5/3} + \left( \frac{2Z}{A} \right)^{5/3} \right]. \end{aligned} \quad (52)$$

Taking

$$\frac{2N}{A} = 1 + \frac{N-Z}{N+Z} \quad \text{and} \quad \frac{2Z}{A} = 1 - \frac{N-Z}{N+Z} \quad . \quad . \quad (53)$$

and developing in powers of the fraction, we obtain the kinetic energy per particle:

$$\frac{E_{\text{kin}}}{A} = +15.7 + 8.6 \left( \frac{N-Z}{N+Z} \right)^2 \quad . \quad . \quad . \quad (54)$$

Thus there appear terms of the same form as those in the formula obtained from the nuclear-energy surface, but with different numerical coefficients. The first term shows that the kinetic energy is exactly equal to the negative of the total energy, as already remarked. The term in  $(N-Z)^2/(N+Z)^2$  also contains a considerable amount of kinetic energy.



## 12. Independent Particle and Shell Models of the Nucleus; Nuclear Moments.

The liquid drop model treated as a Fermi gas accounts in a concise way for many properties of nuclei, just as the treatment of the electron shells as a Fermi gas accounts for many features of atomic structure. However, details such as the mechanical and magnetic moments of nuclei fall outside the province of the drop model. In many cases, the finer details can be deduced from a model in which a single particle is considered and its motion in the schematized potential field of the others is studied. The question of nuclear moments is one that comes particularly within the scope of such considerations.

The experimental fact that the *ee*-nuclei have zero moment (§ 2) indicates that for pairs of like nucleons (proton-proton or neutron-neutron) the angular momenta cancel by setting themselves in the antiparallel position. At the same time, the integral values found for the angular momenta of the four stable *uu*-nuclei at the beginning of the periodic table suggest that the parallel orientation occurs for a proton-neutron pair. Apart from these four instances, it seems natural to ascribe the mechanical as well as the magnetic moment to the unpaired nucleon which is added last. The values of the mechanical moment of the *ue*- or *eu*-nuclei are relatively easy to determine from the hyperfine structure, but it is more difficult to measure the associated magnetic moment. Inasmuch as the nuclear fluid must possess kinetic energy, part of the moment must be the result of its rotation and the rest must come from the spin of the elementary particles themselves. The magnetic moment of a nucleus belonging to a diamagnetic atom or molecule can be determined by a refinement of the atomic beam method (p. 665) developed by Rabi. However, this method cannot be applied to neutrons. Bloch and Alvarez were able to determine the value for the neutron in the following way: A magnetized block of iron is used to align ("polarize") the spins in a beam of neutrons, while a second magnetized block serves as analyser. An alternating magnetic field, applied in the intervening space, disturbs the alignment. This disturbance is greatest when the frequency of the field coincides with the Larmor frequency of the precessing moments.\* The most appropriate unit for expressing nuclear magnetic moments is that corresponding to the revolution of a proton with angular momentum  $h/2\pi$ . Owing to the fact that the proton is 1836 times as massive as the electron, equation (38), p. 464, yields for this quantity the value

$$\mu_N = \mu_B/1836 = (5.048 \pm 0.006) \times 10^{-24} \text{ erg/oersted.}$$

$\mu_N$  is called one nuclear magneton. The experiments referred to above

\* In quantum language, this would occur when the value of  $h\nu$  is equal to the difference in energy of the two orientations in the field.

yield the following values for the magnetic moments of the proton and the neutron:

$$\mu_p = 2.7927\mu_N: \quad \mu_n = -1.9130\mu_N.$$

In each instance the gyromagnetic ratio (p. 671) is double the number above, since each of these particles has spin  $1/2$ . With regard to the signs, the  $+$  sign for the proton indicates that the spin and the magnetic moment must have the same direction ( $+$  charge rotating), while the opposite must be true for the neutron. The sign is determined experimentally from the hyperfine structure.

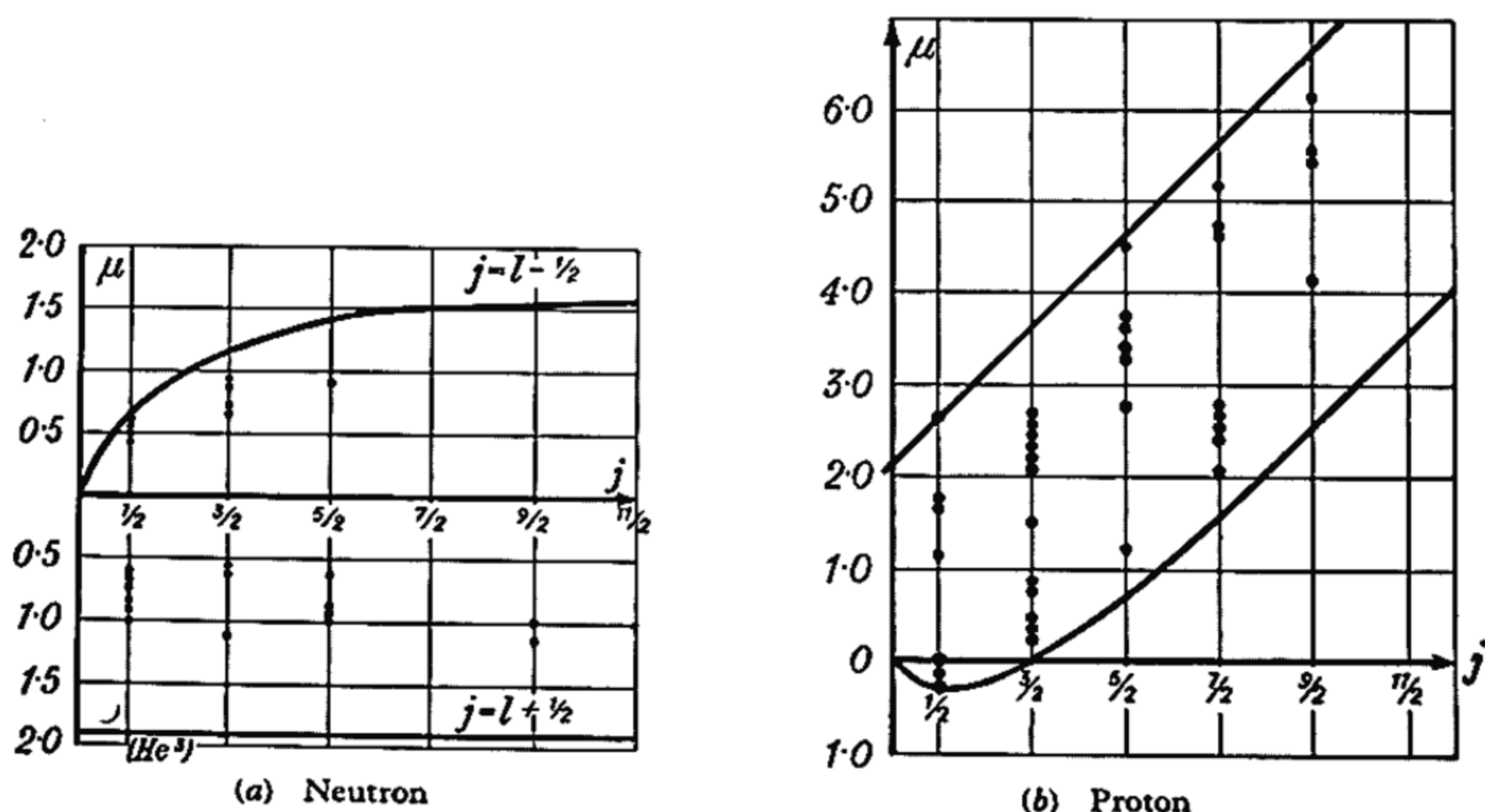


Fig. 7

If there is but one bound nucleon, the resulting magnetic moment may be calculated according to the method of p. 670 by considering the resultant of the orbital moment  $l$  and the spin moment  $s$ , both in units of  $h/2\pi$ , precessing about their resultant,  $j$ . Starting with an arbitrary angle between  $l$  and  $s$ , the cosine is expressed in terms of  $j$ ,  $l$  and  $s$ , with each quantity  $x^2$  replaced by the quantum-mechanical square,  $x(x+1)$ . Finally, the angle is taken to be either  $0$  or  $2\pi$ , the only possible values. Calling the gyromagnetic ratio for the spin  $g_s$ , and leaving the value  $g_l$  for the orbital ratio open, the result is

$$\text{For } j = l + \frac{1}{2}, \quad \mu = g_l l + \frac{1}{2} g_s$$

$$\text{For } j = l - \frac{1}{2}, \quad \mu = g_l l - \frac{1}{2} g_s + \frac{g_s - g_l}{2l + 1}.$$

The orbital  $g$ -factor is set equal to unity for protons; but neutrons, because of the absence of charge, can have no orbital moment and so  $g_l$  must be taken to be zero.

Using the above experimental values of  $g_s$  (5.58 for protons,  $-3.82$  for neutrons), and plotting the values of  $\mu$  as functions of  $j$ , we get the so-called Schmidt-Schüler curves on which the observed moments should fall. Actually, only a few points fall near the curves themselves, most of them scattering over the space between the "Schmidt limits". Taking account of errors of measurement, no point falls definitely outside this region. The reason for the scatter is to be sought in the fact that the unpaired nucleon is not completely independent of the others, as was assumed.

If we wish to compute the energy of excitation or of complete removal of a nucleon following the method used for the electron shells, we must make some sort of assumption regarding the trend of the short-range nuclear forces. A possibility that takes into account the short range of nuclear forces is the square-well potential function, mentioned on p. 770, for which  $V$  is zero beyond a certain distance  $a$  and equal to a constant value  $-V_0$  within this distance. Using this assumption, the simple case of the deuteron will be treated. The Schrödinger equation is

$$\Delta\psi + \frac{8\pi^2 M}{h^2} [E - V(r)]\psi = 0. \quad . \quad . \quad . \quad (55)$$

For the deuteron,  $M$  is equal to the "reduced mass" (p. 651),  $\frac{1}{2}M_H$ . For heavy nuclei,  $M$  can be taken to be that of the last bound nucleon, or  $M_H$ . Since the magnetic moment of the deuteron turns out to be very nearly equal to the sum of the magnetic spin moments of proton and neutron, there is no orbital moment. This means that the quantum number corresponding to orbital motion is zero and the  $\psi$  function is spherically symmetric. If we set

$$\psi = \frac{u(r)}{r} \quad . \quad . \quad . \quad . \quad . \quad . \quad (56)$$

we get a simple equation for  $u$ :

$$\frac{d^2 u}{dr^2} + \frac{8\pi^2 M}{h^2} [E - V(r)]u = 0. \quad . \quad . \quad . \quad (57)$$

In order to avoid difficulties with the algebraic sign, we set the negative binding energy  $E$  equal to  $-W$  and obtain

$$\left. \begin{aligned} \frac{d^2 u}{dr^2} + \frac{8\pi^2 M}{h^2} (V_0 - W)u &= 0 \quad \text{for } r < a; \\ \frac{d^2 u}{dr^2} - \frac{8\pi^2 M}{h^2} W u &= 0 \quad \text{for } r > a. \end{aligned} \right\} \quad . \quad . \quad (58)$$





For higher quantum states the expression for  $\psi$  also involves surface harmonics multiplied by a function of  $r$ . As for atoms, an orbital quantum number  $l$  is introduced in order to describe the energy states. A lengthy calculation yields the following term sequence for the case of the square-well potential functions:

1s	2p	3d	2s	4f	3p	5g	4d	6h	3s
2	6	10	2	14	6	18	10	22	2

The numbers in the lower line are the maximum numbers of nucleons that can reside in each state, taking account of the Pauli Principle as applied to the two possible orientations for the spin. These numbers hold for protons as well as for neutrons, since these are really different particles which are distinguishable by their charge.

The assumption of a sharp-edged square well is obviously a bold idealization of the actual situation. A more realistic assumption which is at the same time amenable to calculation is the oscillatory potential

$$V = V_0 - \frac{1}{2}M\omega_0^2 r^2. \quad (64)$$

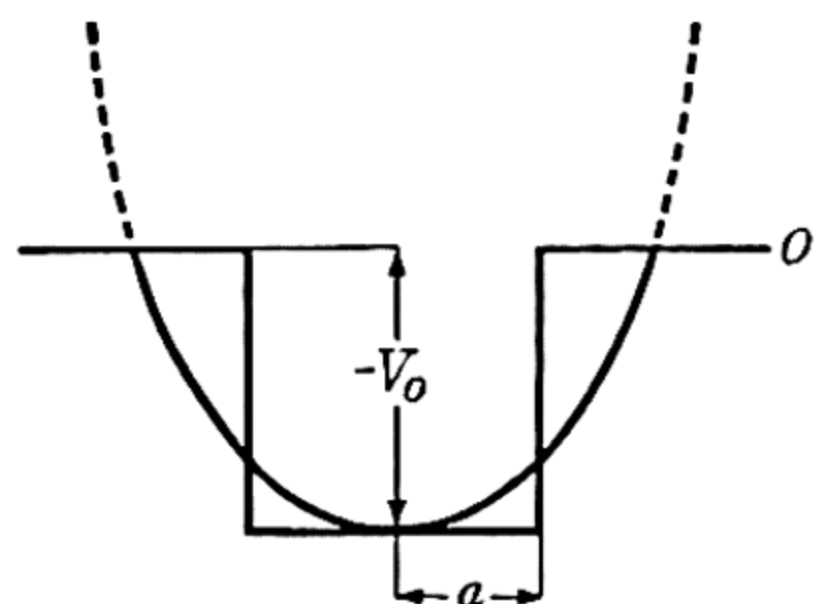


Fig. 8

The depth of the well is characterized by the constant  $V_0$ , while the width is the opening in the parabola, which depends

on the value of  $\omega_0$  (fig. 8). While this representation is far from the actual situation for large values of  $r$ , this is of little consequence because the particle is seldom found at these large distances. In this case the term sequence and particle numbers turn out to be:

0s	1p	2d	2s	3f	3p	4g	4d	4s	5h	5f	5p
2	6	10	2	14	6	18	10	2	22	14	6

A somewhat better approach to the truth is obtained by averaging both potentials by taking the mean of each pair of closest-lying levels, which are given by the same value of  $l$ .

The spin, which has the effect of bringing about a further splitting of the energy levels, has not been taken into account up to this point. The spin-orbital interaction is greatest here, so that the situation is analogous to the  $jj$ -coupling of the electron shells rather than to the  $LS$  coupling to the other nucleons, which was taken into account in the foregoing treatment by means of the nuclear potential. As in the case of the electron shells, the splitting must increase with increasing  $l$ . If a direct proportionality between the splitting and  $l$  is assumed, the partial levels of one term occasionally fall between those of another and a new grouping results which corresponds neither to that of the square well



nor the oscillating potential function. There are gaps of considerable size in the energy levels following these numbers of protons or neutrons:

2      8      20      (28)      (40)      50      82      126

A number in parentheses means that the gap following it is present but not as marked as for the others.

What do these gaps signify? In the theory of electron shell structure, the closed shell of the noble gases is characterized by the fact that the energy levels nearest the ground state are especially far above this state; e.g. the electron is particularly tightly bound. For the alkali metal immediately following, however, the electron is especially loosely bound. In exactly the same way, a nucleus whose number of neutrons or protons corresponds to one of the above so-called "magic numbers" is particularly stable and symmetric. This affects all nuclear properties. For example, we note the large number of isotopes of Sn ( $Z = 50$ ). The quadrupole moment of such nuclei is very small, increasing with the addition of each nucleon (cf. fig. 5). Thus the magic numbers correspond to closed shells, just as do the numbers of electrons in the outer shells of the noble gases. It must be remarked that 50 neutrons as well as 50 protons gives a particularly stable nucleus. This makes plausible the unusual stability of "doubly magic" nuclei, of which  ${}^4_2\text{He}$  is the first.

### 13. Beta Transformations and the Neutrino Hypothesis.

The established fact that the energy distribution of the electrons emitted in  $\beta$ -transformations is continuous, is in sharp contradiction to all other observations which show that the nucleus, like the electronic shell, has a number of definite quantum states of energy. If a number of nuclei were to disintegrate with the emission of electrons only, the new nuclei would have a continuously varying content of energy. How can such nuclei, which occur in the radioactive series, emit  $\alpha$ -rays all having the same sharp range? Here we have the choice between only two courses: give up the conservation of energy for  $\beta$ -transformations, or assume that a particle is emitted which escapes observation. So long as there is no other evidence for the latter assumption, it amounts of course to no more than a plausible way of getting around the former disagreeable assumption. There is, however, other evidence that another particle is emitted. We saw (p. 744) that, according to observation, nuclei of odd atomic weight have moments which are multiples of  $1/2$ . In  $\beta$ -disintegration the atomic weight is not altered but the moment changes by a multiple of  $1/2$ , since the electron spin is  $1/2$  and only a whole-number orbital moment can be added to this. This difficulty is surmounted by giving the hypothetical particle, like most of the stable particles, a spin of  $1/2$ . Its mass must be much



smaller than that of the electron, presumably less than  $1/5000$  the electron mass. In nuclear reactions where such particles must be assumed to be produced (e.g. the decay of the neutron) the charges of the remaining products are found to balance. This, together with the fact that the new particle produced no observable track, indicates that it has zero charge. This particle, whose existence had to be assumed in order to retain the conservation of energy in a number of observed nuclear reactions, was named the *neutrino*.

Recent cloud chamber observations of  $\beta$ -decay, undertaken for this purpose, show that the conservation of momentum is not fulfilled for the emitted electron and recoil nucleus alone. More recent experiments for detecting possible neutrino capture by protons now have yielded direct evidence for the existence of the neutrino.

A further circumstance to be considered is the absence of any annihilation radiation in K-capture (§ 7). Meitner performed an experiment which showed that neutrinos are not simply  $\gamma$ -radiation. The heat developed in a  $\beta$ -reaction was measured in a calorimeter which was arranged to trap any  $\gamma$ -radiation which might have been produced. Had the heat been produced by  $\gamma$ -rays, its amount would have corresponded to the maximum energy of the electrons. It was found, however, to correspond to the *average* energy of the electrons.

Neutrinos have such small interaction with matter that they appear to pass right through the solid earth without appreciable absorption. The question of what happens eventually to their energy is a very disturbing one, particularly since it is far from trivial in amount. It is estimated, for example, that of the total energy production of a uranium reactor, some 2000 kilowatts is carried off by neutrinos.

#### 14. Mesons.

Besides the new particles already mentioned—neutron, positron and neutrino—there exists a group of particles whose masses lie between that of the electron and that of the proton. These particles, called *mesons*, occur both with positive and with negative charge, and certain ones in the neutral state as well. The first indication of the existence of mesons was obtained in cloud-chamber work with cosmic rays. Tracks were found that were definitely thinner than those of protons and they were formerly ascribed to electrons. But according to the photographs the loss of momentum of these particles in passing through a lead plate is only one-tenth of that to be expected from an electron. This difficulty can be avoided by assuming that the particles have a mass of about 220 times that of the electron. Yukawa had already deduced the existence of such a particle from considerations which at first sight seem more formal. For the electrostatic field we have Laplace's equation  $\Delta\phi = 0$ . Since nuclear forces must fall away





### 15. Brief Survey of Our Knowledge concerning Cosmic Rays.

As far as the amount of converted energy is concerned, chemical processes may be characterized as the "physics of electron volts", nuclear transformations as "Mev physics" and cosmic-radiation phenomena as "million Mev physics".\* It is to be expected that an increase of many orders of magnitude in the energy involved brings with it entirely new phenomena. The first thing that had to be done in order to explain the nature of cosmic radiation was to distinguish the effects of the primary rays from those of the secondary. The effects of the primaries must increase monotonically with altitude, reaching a constant value beyond the atmosphere. The effects of the secondaries, on the other hand, must attain a maximum at a certain height; for they originate in the air itself and the primary radiation which causes them will be progressively absorbed.

Apart from a small contribution of secondary particles scattered from below, the primary rays can be studied in almost pure form at great altitudes. Experiments have been carried out in aircraft up to about 10 miles, recording apparatus and photographic plates have been taken to heights of about 20 miles in unmanned balloons, and high-altitude rockets have carried instruments to over a hundred miles. The principal effect of the primary particles is the production of single, very energetic collisions which show up on a photographic plate as "stars" in which many nucleon and  $\pi$ -meson tracks radiate from the point of impact. The number of ion pairs simultaneously produced by these particles is sufficient to actuate an electrometer of high sensitivity.

At the present time, the most important means of investigating the primaries is the photographic-plate technique. Here the tracks are produced by fragments of silver or bromine atoms which have been disintegrated by the intrusion of a primary particle. What is the nature of these particles? They must be charged, since their distribution is found to depend on the geomagnetic latitude of the place of observation. To-day it can be considered well established that the majority of the primary particles are protons. There are also heavier particles present, ranging up as far as Fe nuclei,  $\alpha$ -particles being especially numerous. Because of their greater range, the protons predominate at greater depths.

At the earth's surface and at moderate altitudes only few primary particles are observed, most of the rays being a mixture of quite diverse secondary particles. Although very energetic secondary nucleons and  $\pi^-$ -mesons are able to produce small stars, these particles constitute

\* In American publications the abbreviation Bev is sometimes used to denote a unit of  $10^9$  electron volts. It must be remembered that the term "billion" stands for  $10^9$  in America and for  $10^{12}$  in Britain.



only a small part of the secondary radiation.\* It has been known for a long time that the radiations detected at lower levels in the atmosphere consist of two components—a penetrating (hard) component and a softer one. The separation of the components and the determination of their properties have been the object of an extremely large amount of experimental and theoretical work. It has been shown that the hard component consists of  $\mu$ -mesons while the soft component is made up of less energetic electrons and  $\gamma$ -photons occurring in the form of extended showers stretching over distances often exceeding hundreds of meters.

The  $\mu$ -mesons, formed by the disintegration of  $\pi$ -mesons, have a lifetime long enough to permit them to penetrate deep into the earth. Their ionizing power is so slight compared with their energy that a particle having an energy of  $10^{11}$  ev is able to penetrate 500 metres of water. The electrons and  $\gamma$ -photons begin to appear even at the very top of the atmosphere. Here the short-lived neutral  $\pi$ -mesons decay into two  $\gamma$ -photons. By pair production and Compton processes these give rise to fast electrons. The electrons, in turn, produce  $\gamma$ -photons when retarded, and so the process continues. The result is a cascade shower in which the available energy is rapidly shared among a large number of particles of low penetrating power.

Special triggering arrangements are needed to catch the beginning of a large shower in a cloud chamber experiment. If the point of origin happens to be located in the material in a large chamber, the electrometer will show an abrupt impulse as for an explosive shower (see above). Usually only tracks of individual electrons or positrons are seen in the cloud chamber, and sometimes  $\mu$ -meson tracks as well.

The biggest question of all is that of the origin of cosmic radiation. At first it was thought that the rays were due to some mysterious instance of conversion of matter into energy. This is known to be false, if only for the reason that the annihilation of even the most massive atom would not be enough to account for the energies of up to perhaps  $10^{17}$  ev that are encountered. In the light of present experience with particles accelerated in the laboratory, it seems plausible that the cosmic particles obtain their energy from stray magnetic fields in space. The particles may in part originate in the sun. The magnetic fields, because of their enormous extent in space, need not be strong, but might be more concentrated in the neighbourhood of certain stars, especially those showing strong emission of radio waves. It is still very much an open question as to which of the possible mechanisms actually obtain. The fact that the observed intensity is the same in all directions is in accord with such theories, for the particles are probably under way for thousands of years before reaching the earth.

\* The  $\pi^+$ -mesons are repelled by the nuclear charge and so do not ordinarily lead to disintegrations.



## **PART VIII**

### **SELECTED TOPICS FROM SEVERAL FIELDS OF TECHNICAL IMPORTANCE**





## CHAPTER XLIII

### FURTHER MATTERS CONNECTED WITH THE GEOMETRIC OPTICS OF LIGHT AND OF ELECTRONS

#### 1. Condition for Absence of Distortion and its Relation to the Sine Condition.

On p. 406 we deduced the condition governing the point-for-point imaging of an element of area normal to the axis with the provision that axial points be imaged without distortion by means of wide pencils of rays. Inasmuch as the squares of the linear dimensions of the area were neglected in this derivation, the resulting *sine condition* guarantees only the imaging of a small area, e.g. the field of view of a microscope. But if requirements in the quality of the image near the boundaries are less stringent, the range of validity of the result may be considerably extended. We wish now to consider the other extreme—the imaging of a large field of view by means of narrow pencils. The limiting case is that of the pinhole camera, which uses no optical unit (such as a lens) at all. The sharpness of the image depends solely on the size of the aperture. We need not use such narrow bundles of rays if we use an optical system, but they must still be fairly small if other errors such as astigmatism (p. 409) are to be minimized. It is a challenge to the skill of the optical designer to keep these other defects small so that good images result even when a large field of view and wide pencils are employed. We shall not concern ourselves with these other errors.

Let the cross-section of the entering bundle of rays be limited by a *diaphragm*  $AB$ , fig. 1, assumed to be placed in front of a lens system  $S$ , indicated by the two curved surfaces. In this case  $AB$  is also called the *entrance pupil*. If there happen to be lenses on the side toward the object, the entrance pupil is found by locating the image of the diaphragm as formed by these lenses. The reason is that a ray passing through the edge of the entrance pupil must then pass through the edge of the diaphragm, since entrance pupil and aperture correspond to each other. In the same way, the boundaries of the pencil of rays on the image side are determined by locating the image of the diaphragm as formed by any lenses lying on this side of it. This image of the diaphragm opening is called the *exit pupil* and is represented by  $A'B'$  in the figure.

Further, the rays drawn from individual points of the object or of the image to the centre of the corresponding pupil ( $O$  or  $O'$  in the sketch) are called *principal rays*. These rays are taken to represent the pencils traversing the pupils from each of the various object points. Let  $P, Q, R$  represent three points located on a perpendicular to the optical axis, and let  $P', Q', R'$  be their respective images. What condition must hold in order that object and image be similar? Obviously we must have  $P'R' : Q'R' = PR : QR$ . In terms of the labelling used in the diagram, this means

$$\tan u_1' / \tan u_2' = \tan u_1 / \tan u_2,$$

or the condition for similarity (distortionless imaging) becomes

$$\frac{\tan u_i}{\tan u_i'} = \text{const.} \quad \cdot \cdot \cdot \cdot \cdot \quad (1)$$

If we compare this with the sine condition, the two requirements seem at first sight incompatible. However, at least within certain

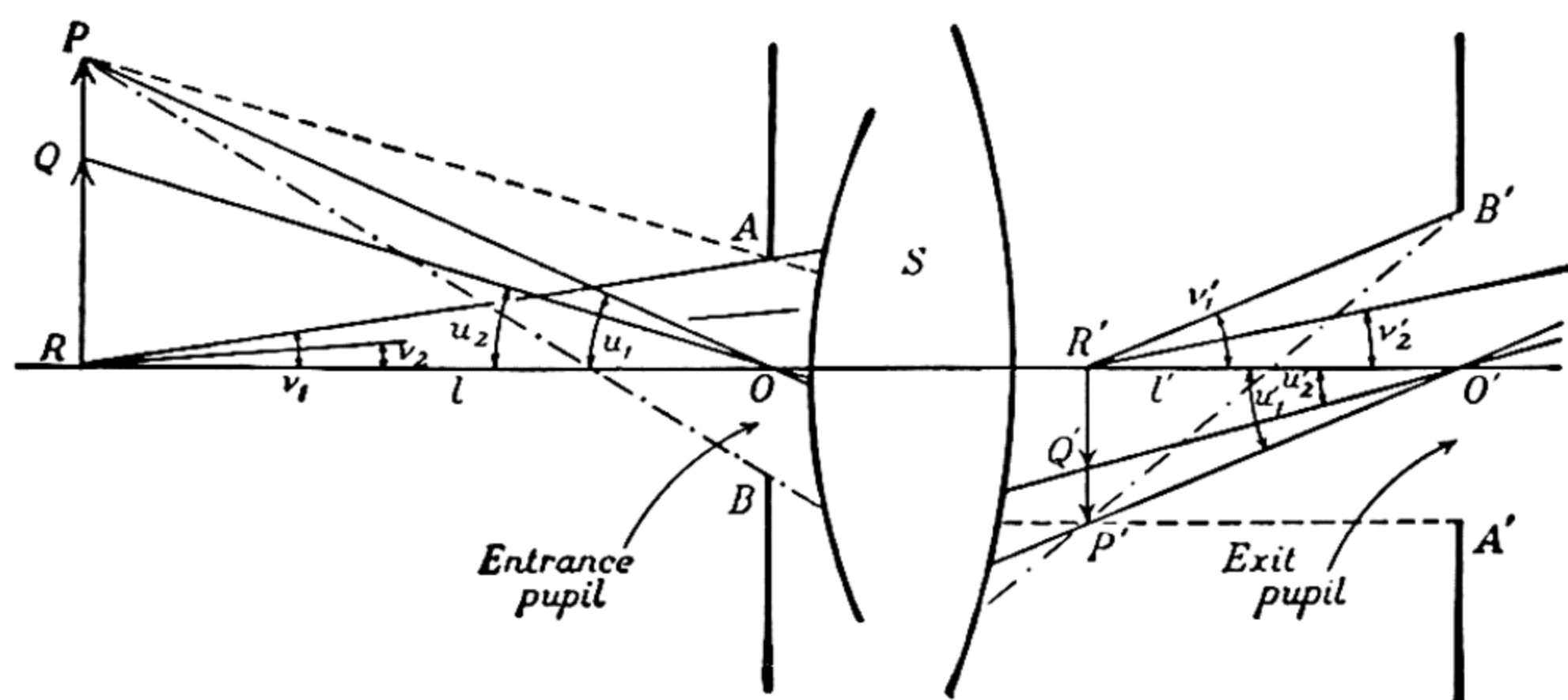


Fig. 1

limits, this is not so. The sine condition refers to the angles between the axis and the various rays drawn from an axial point  $R$  through the several points of the entrance pupil, as indicated by  $v_1, v_2$ , etc., in fig. 1. In other words, the system  $S$  must be figured so that it fulfils the sine condition for the plane  $PQR$  as a whole and the tangent condition for the principal rays. The realization of good definition is the first consideration, while in many applications some distortion can be tolerated. For example, if the system is so corrected that the sine condition is fulfilled for the imaging of the entrance and exit pupils, there will still be distortion in the final image. It is easy to compute how a rectangular network placed in the plane  $P'Q'R'$  will be imaged when, in place of the tangent condition, the sine condition is satisfied. Re-



versing affairs now, and placing the *distorted* grid at  $PQR$ , its image at  $P'Q'R'$  will be rectangular. This was used by Abbe as a sensitive test of the sine correction of microscope objectives. He placed the entrance pupil  $AB$ —a small hole—in the object plane of the objective to be tested, and noted whether the image formed at  $P'Q'R'$ , the plane for which the figure was computed, was free of distortion. Fig. 2 shows one quadrant of such a test figure.

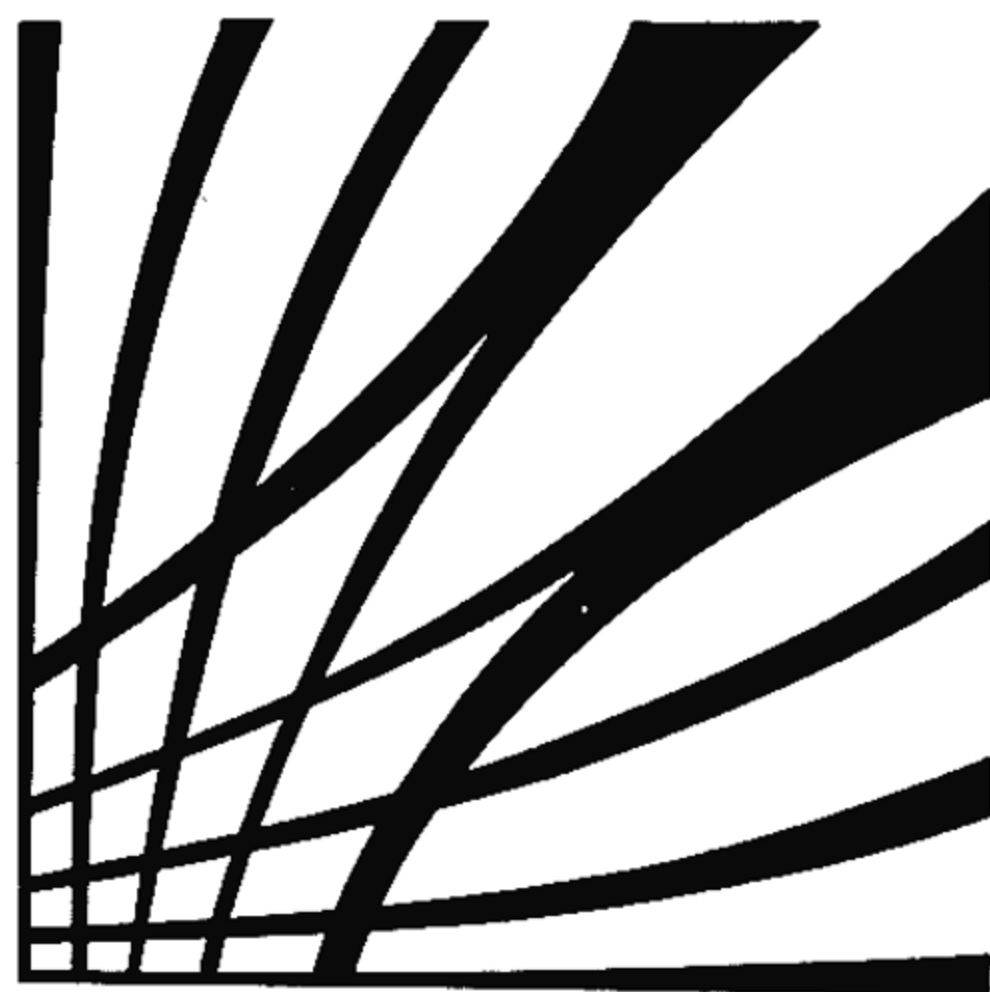


Fig. 2

## 2. Focal Length of an Electron Lens.

The parallelism between ray optics and the mechanics of particles explained on p. 693 shows that in representing the orbit of an electron in an electric field by a ray of light in a homogeneous medium we must think of the medium as having an index of refraction proportional to the speed of the electron. This follows from a comparison of equations (15) and (16). As long as we confine attention to *geometric* electron-optics, the factor of proportionality is immaterial and equation (23) is suitable for a wave-mechanical treatment. If the electrons owe their velocities only to electric fields, i.e. if their initial thermal motion can be neglected, then  $n$  may be set equal to the square root of the potential difference  $V$  that exists between the region considered and the cathode. As long as only rays near the axis are considered, a rotationally symmetric electric field (fig. 3) corresponds to a series of centred spherical surfaces as already treated on p. 405. The sole difference between this case and ordinary optics is that here the index of refraction changes continuously, whereas in optics spherical surfaces usually separate regions in each of which the index is constant but different from one region to the next. This discontinuity does not introduce the slightest difficulty as long as the distance between shells can be neglected. This,

of course, makes the result a fairly rough approximation. Successive application of equation (25), p. 405, to a series of  $k$  layers bounded by centred spheres gives, for the *power*,

$$\phi' = \frac{1}{n_{k+1}} \sum_1^{k+1} n_i \phi'_i. \quad . \quad . \quad . \quad . \quad . \quad (2)$$

On the other hand, from equation (21), p. 405, the power of a spherical surface of radius  $\rho$  and discontinuity of index amounting to  $\Delta n_i$  is

$$\phi_i = \frac{\Delta n_i}{n_i \rho}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

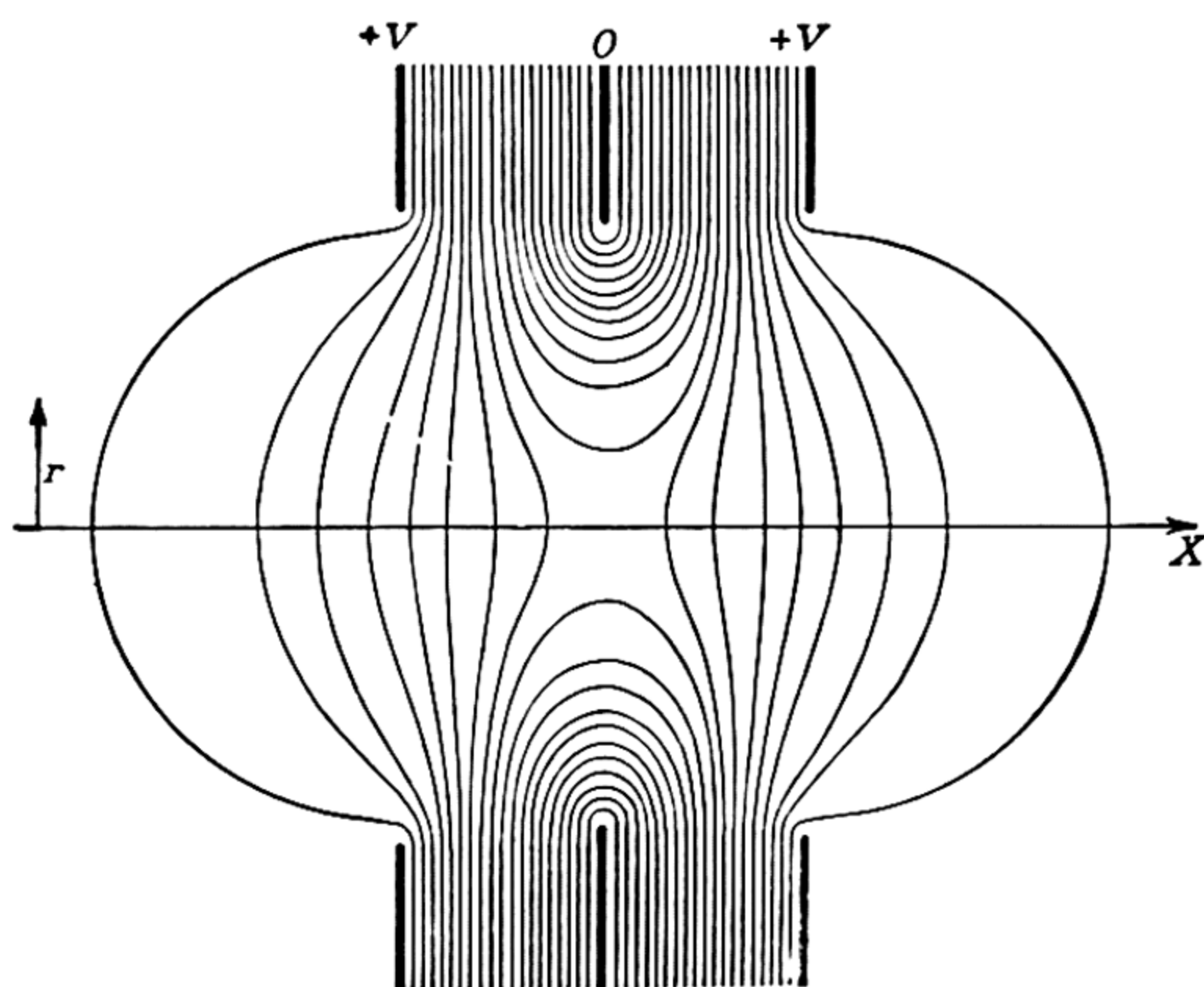


Fig. 3

If  $n$  changes continuously along the axis,

$$dn = \frac{dn}{dx} dx \quad . \quad . \quad . \quad . \quad . \quad (4)$$

and

$$\phi' = \frac{1}{n'} \int \frac{1}{\rho} \frac{dn}{dx} dx. \quad . \quad . \quad . \quad . \quad . \quad (5)$$

In the electrical case,  $n = \sqrt{V}$ , and so

$$\phi = \frac{1}{\sqrt{V'}} \int_{-\infty}^{+\infty} \frac{1}{\rho} \frac{d\sqrt{V}}{dx} dx = \frac{1}{2\sqrt{V'}} \int_{-\infty}^{+\infty} \frac{1}{\rho} \frac{dV}{dx} \frac{1}{\sqrt{V}} dx. \quad (6)$$

The integration may be performed between the limits  $-\infty$  and  $+\infty$  because  $dV/dx$  vanishes in the field-free space and there is no con-

tribution except where  $dV/dx$  differs from zero. The quantity  $V'$  denotes the constant potential that exists at a great distance beyond the lens.

The radii of curvature of the equipotential surfaces may be found in terms of the change of  $V$  measured along the axis by using differential geometry and the potential equation  $\Delta V = 0$ . This fact brings out a fundamental difference between ordinary optics and electron optics which complicates the process of computing electron lenses. In the case of light, the curvatures and indices of refraction of lenses can be selected independently, and so by skilful choice the lens errors can be compensated. In electron optics, however, the radii are already determined by the change of index along the axis. The result of the somewhat cumbersome calculation, which may be transformed by partial integration, takes the following forms, since  $dV/dx$  vanishes at great distances:

$$\begin{aligned}\phi' &= \frac{1}{4\sqrt{V'}} \int_{-\infty}^{+\infty} \frac{d^2V}{dx^2} \frac{1}{\sqrt{V}} dx = \frac{1}{8\sqrt{V'}} \int_{-\infty}^{+\infty} \left(\frac{dV}{dx}\right)^2 V^{-3/2} dx \\ &= \frac{1}{8\sqrt{V'}} \int_{-\infty}^{+\infty} \frac{E^2}{V^{3/2}} dx. \quad . . . . . (7)\end{aligned}$$

As we can see by considering fig. 3, the repeated neglect of the distance between lenses must result in a considerable total error, but the optical computation taking these distances into account becomes very troublesome. Generally, in electron optics, the calculation of electron paths leads more quickly to the result than does the formal comparison with ordinary optics. Moreover, the former has the advantage that all the required equations are already at hand. For this reason we shall compute the focal length only in this way. As an approximation for narrow pencils we equate the projection of the orbital velocity on the axis to this velocity itself. Then

$$\frac{dx}{dt} = v_x = v = \sqrt{\left(\frac{2e}{m} V\right)}, \quad . . . . . (8)$$

where  $V$  is the potential referred to the cathode.

Let the radial component of the field at a distance  $r$  from the axis be denoted by  $E_r$ . Then, since

$$\frac{dr}{dt} = \frac{dr}{dx} \frac{dx}{dt} = v \frac{dr}{dx},$$

we have

$$\frac{d^2r}{dt^2} = v \frac{d}{dx} \left( v \frac{dr}{dx} \right) = -\frac{e}{m} E_r. \quad . . . . . (9)$$



In a charge-free space,  $E_r$  can be expressed in terms of  $E_x$ , the component of the field along the axis. Imagine a cylinder of radius  $r$  and short length  $l$  described about the axis, and calculate the surface integral of the field over this cylinder. It may usually be assumed that there is no charge within the cylinder since the space charge due to the electrons is negligible. Then the surface integral is

$$2\pi r l E_r + \pi r^2 l \frac{dE_x}{dx} = 0 \quad \text{or} \quad E_r = -\frac{r}{2} \frac{dE_x}{dx} = \frac{r}{2} \frac{d^2V}{dx^2}. \quad (10)$$

The differential equation of the electron orbit then becomes

$$\sqrt{V} \frac{d}{dx} \left( \sqrt{V} \frac{dr}{dx} \right) = -\frac{r}{4} \frac{d^2V}{dx^2}. \quad (11)$$

Examination of this equation shows that:

1. The specific charge  $e/m$  does not appear, so that the result holds for charged particles of any kind.
2. The equation is homogeneous in  $V$ , so that changing the potentials of all electrodes by the same amount leaves everything the same. This shows why an electrostatic electron microscope is not sensitive to variations in voltage.
3. The equation is homogeneous in  $r$  and  $x$ . Thus, if all the dimensions of an electron microscope are changed by a given factor, the magnification remains unchanged.

In order to integrate the equation of the orbit, write it in the form

$$V(x) \frac{d^2r}{dx^2} + \frac{1}{2} \frac{dV(x)}{dx} \frac{dr}{dx} + \frac{1}{4} \frac{d^2V(x)}{dx^2} r = 0. \quad (12)$$

This shows that we are dealing with a second-order linear differential equation whose coefficients are functions of  $x$  and may be determined by measurement or by calculation. We can bring the equation into simpler form by putting

$$r = R(x)f(V), \quad (13)$$

determining  $f(V)$  so that the term in  $dr/dx$  vanishes. The calculation readily gives

$$f(V) = V^{-1/4} \quad (14)$$

and 
$$\frac{d^2R}{dx^2} = -\frac{3}{16V^2} \left( \frac{dV}{dx} \right)^2 R = F(x)R. \quad (15)$$

The integration may be carried out by successive approximations, as was done in the case of equation (36), p. 101.

In order to compute the focal length, consider a ray in the object space, parallel to the axis and located a distance  $a$  from the axis. Its

intersection with the axis in the image space determines the principal focal point  $F'$ . From the definition of focal length (p. 399),

$$a = -f' \left( \frac{dr}{dx} \right)_{F'} \quad \text{or} \quad \phi' = \frac{\left( \frac{dr}{dx} \right)_{F'}}{a}. \quad . \quad . \quad . \quad (16)$$

Since the focus lies in a region of constant potential  $V'$ , (13) yields

$$\left( \frac{dr}{dx} \right)_{F'} = \left( \frac{dR}{dx} \right)_{F'} V'^{-\frac{1}{2}}. \quad . \quad . \quad . \quad . \quad (17)$$

In front of the lens we have the constant potential  $V_0$ , and so

$$a = R_0 V_0^{-\frac{1}{2}}. \quad . \quad . \quad . \quad . \quad . \quad (18)$$

If we now wish to compute  $(dR/dx)_{F'}$  from equation (15), we can, as a first approximation, take the value  $R = aV_0^{\frac{1}{2}}$  which holds to the left of the lens. By integration we then get

$$\left( \frac{dR}{dx} \right)_{F'} = \left( \frac{dR}{dx} \right)_{\infty} = -\frac{3}{16} a V_0^{\frac{1}{2}} \int_{-\infty}^{+\infty} \left( \frac{dV}{dx} \right)^2 \frac{1}{V^2} dx, \quad (19)$$

where the upper limit may be taken to be infinity because the focus is in a field-free region. The power of the system then becomes

$$\phi' = \frac{3}{16} \frac{V_0^{\frac{1}{2}}}{V'^{\frac{1}{2}}} \int_{-\infty}^{+\infty} \left( \frac{dV}{dx} \right)^2 \frac{1}{V^2} dx. \quad . \quad . \quad . \quad . \quad (20)$$

For a single lens the potential has the same value on both sides, and a better expression for the power of a single lens is obtained by replacing  $dV/dx$  by  $\mathbf{E}$ :

$$\phi' = \frac{3}{16} \int_{-\infty}^{+\infty} \frac{E^2}{V^2} dx. \quad . \quad . \quad . \quad . \quad (21)$$

Similarly, the power of a lens formed by a short, rotationally symmetric magnetic field  $\mathbf{H}$  is found to be

$$\phi' = \frac{e}{8mV} \int_{-\infty}^{+\infty} H^2 dx, \quad . \quad . \quad . \quad . \quad (22)$$

$V$  being the potential through which the electrons have fallen.

## CHAPTER XLIV

### PIEZOELECTRICITY AND ITS APPLICATIONS

#### 1. The Phenomenon of Piezoelectricity.

When an ionic crystal is deformed an electrical volume-polarization results, and this, according to the discussion on p. 281, makes itself known by the appearance of charges on the surface. In a rough way, the occurrence of this piezoelectric polarization may be pictured by thinking of a regular tetrahedron whose centre is occupied by an ion

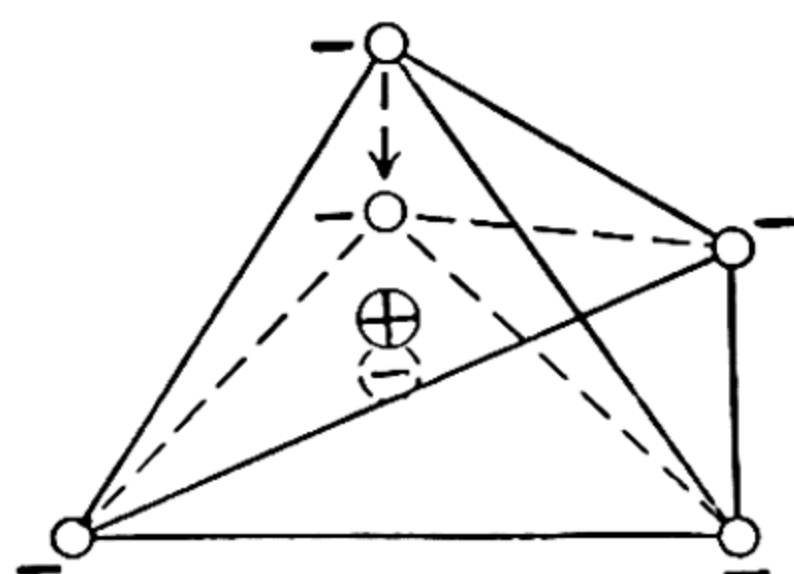


Fig. 1

carrying four positive charges, while each corner is occupied by a singly charged negative ion (fig. 1). Inasmuch as the centre of gravity of the negative charge coincides with the position of the positive charge, the structure as a whole has no dipole moment. If, now, the central ion is held in position while the top point is pushed downward, a dipole moment will obviously arise. In simple crystal lattices such as NaCl there is no resul-

tant dipole moment per unit volume, since the symmetry of the lattice insures a mutual compensation. The fundamental requirement for the appearance of piezoelectricity is the presence of what is called a polar axis—an axis such that the plane normal to it is *not* a plane of symmetry of the crystal. The relation between the deformation and the resulting polarization is linear. We can therefore write for the components of the polarization vector:

$$\left. \begin{aligned} P_x &= \gamma_{11}e_{11} + \gamma_{12}e_{22} + \gamma_{13}e_{33} + \gamma_{14}e_{23} + \gamma_{15}e_{31} + \gamma_{16}e_{12}, \\ P_y &= \gamma_{21}e_{11} + \gamma_{22}e_{22} + \gamma_{23}e_{33} + \gamma_{24}e_{23} + \gamma_{25}e_{31} + \gamma_{26}e_{12}, \\ P_z &= \gamma_{31}e_{11} + \gamma_{32}e_{22} + \gamma_{33}e_{33} + \gamma_{34}e_{23} + \gamma_{35}e_{31} + \gamma_{36}e_{12}. \end{aligned} \right\} \quad (1)$$

The  $e_{ik}$  are the components of the strain tensor as defined on p. 164; the coefficients  $\gamma_{ik}$  are called the *piezoelectric constants*. Depending upon the degree of symmetry of the crystal, the actual number of constants is reduced. For quartz—the most important of such crystals in practice—the scheme of coefficients referred to a co-ordinate system



conforming to the axes of symmetry (fig. 2) has the following appearance:

$$\left. \begin{array}{cccccc} \gamma_{11} & -\gamma_{11} & 0 & \gamma_{14} & 0 & 0 \\ 0 & 0 & 0 & 0 & -\gamma_{14} & -\gamma_{11} \\ 0 & 0 & 0 & 0 & 0 & 0 \end{array} \right\} \cdot \quad (2)$$

In the place of the components of the strain tensor one may introduce those of the stress tensor.\* In a crystal, however, the relation between the two tensors is much more complex than in isotropic bodies, where two elastic moduli—for example,  $E$  and  $\sigma$ —give the com-

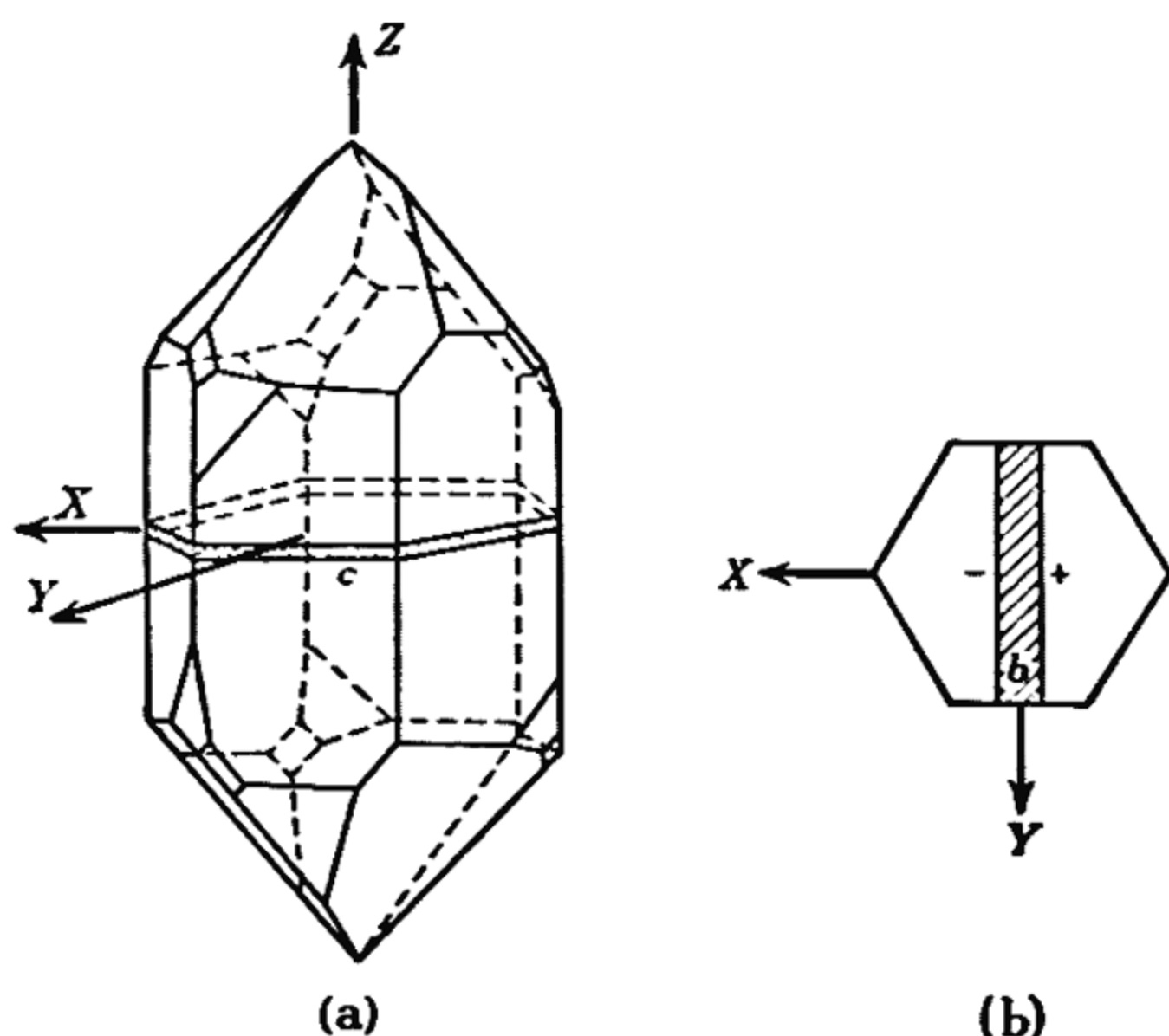


Fig. 2

plete connexion. In the general case, the symmetric scheme of coefficients relating the six components of the stress tensor with the six coefficients of the strain tensor contains 21 different constants. Expressing the  $e_{ik}$  in terms of the  $P_{ik}$  by means of these relations, we get:

$$\left. \begin{array}{l} -P_x = \delta_{11}P_{11} + \delta_{12}P_{22} + \delta_{13}P_{33} + \delta_{14}P_{23} + \delta_{15}P_{31} + \delta_{16}P_{12}, \\ -P_y = \delta_{21}P_{11} + \delta_{22}P_{22} + \delta_{23}P_{33} + \delta_{24}P_{23} + \delta_{25}P_{31} + \delta_{26}P_{12}, \\ -P_z = \delta_{31}P_{11} + \delta_{32}P_{22} + \delta_{33}P_{33} + \delta_{34}P_{23} + \delta_{35}P_{31} + \delta_{36}P_{12}. \end{array} \right\} \quad (3)$$

The  $\delta_{ik}$ , which may be computed from the  $\gamma_{ik}$  and the elastic constants, are called the *piezoelectric moduli*.

\* Since the discovery of piezoelectricity by J. and P. Curie it has been the practice to work with the pressures rather than with their opposites—the tensions. In what follows, the  $P_x$  represent the components of the pressure tensor, and this is the reason for introducing the negative signs in (3), following.

Reciprocally, if an electric field is applied to a piezocrystal, the resulting polarization causes elastic stresses and deformation. When the components of the strain tensor are expressed in terms of the components of the field strength, the coefficients  $\delta_{ik}$  appear. The deduction is similar to that of equation (56), p. 176. We calculate the expression for the energy of the crystal in the electric field. The results for the components of the strain tensor are

$$\left. \begin{aligned} e_{11} &= \delta_{11} E_x + \delta_{21} E_y + \delta_{31} E_z, & e_{23} &= \delta_{11} E_x + \delta_{21} E_y + \delta_{31} E_z, \\ e_{22} &= \delta_{12} E_x + \delta_{22} E_y + \delta_{32} E_z, & e_{31} &= \delta_{15} E_x + \delta_{25} E_y + \delta_{35} E_z, \\ e_{33} &= \delta_{13} E_x + \delta_{23} E_y + \delta_{33} E_z, & e_{12} &= \delta_{16} E_x + \delta_{26} E_y + \delta_{36} E_z. \end{aligned} \right\} \quad (4)$$

For the negative pressure components a scheme analogous to (4) is obtained. The quantities  $\gamma_{ik}$  appear as coefficients of the  $E_i$ ; for quartz we have:

$$\left. \begin{aligned} -P_{11} &= \gamma_{11} E_x, & -P_{23} &= \gamma_{14} E_x, & e_{11} &= \delta_{11} E_x, & e_{23} &= \delta_{11} E_x, \\ P_{22} &= \gamma_{11} E_x, & P_{31} &= \gamma_{14} E_y, & e_{22} &= -\delta_{11} E_x, & e_{31} &= -\delta_{14} E_y, \\ P_{33} &= 0, & P_{12} &= \gamma_{11} E_y, & e_{33} &= 0, & e_{12} &= -2\delta_{11} E_y. \end{aligned} \right\} \quad (5)$$

The position of the co-ordinate system underlying these relations is that of fig. 2. For quartz, the numerical values in C.G.S. units (algebraic signs for right-handed quartz) are:

$$\begin{aligned} \gamma_{11} &= 4.77 \times 10^4, & \delta_{11} &= 3.36 \times 10^{-8}, \\ \gamma_{14} &= 1.23 \times 10^4, & \delta_{14} &= -1.69 \times 10^{-8}. \end{aligned}$$

## 2. Application of Piezoelectricity to the Stabilization of Oscillating Circuits.

The most important application of piezoelectricity is based on the resonance between the frequency of an oscillating circuit and the mechanical vibration of a bar or plate cut from a piezocrystal. Let us consider the simplest case of a bar-like piece cut from a crystal as shown in fig. 2. Suppose that the electric potential is applied to the long sides of the bar, so that the field is in the direction of the  $x$  axis. As a result there will be changes in length in the  $y$  direction and at resonance the bar will experience strong longitudinal vibrations. We must first compute the natural vibration of a bar of length  $l$ . This can be done in a way quite analogous to the calculation of the vibration of a string carried out on p. 183. Here again we can treat the natural vibration as a standing wave resulting from the superposition of two waves travelling in opposite directions. The speed of propagation is given (p. 183) by the square root of the ratio of the elastic constant to the density. For the crystal, we must use the elastic modulus correspond-

ing to the direction of the length of the bar. The boundary conditions here are just the opposite of those obtaining for the string. The greatest amplitude occurs at the free ends of the bar. This result may be deduced formally, but is also intuitively evident, since at the ends the motion of the bar is completely unhindered, whereas in the case of the vibrating string the opposite is true. Thus the fundamental vibration of the bar must have a loop at each end and a node at the centre, so that—as for the string—the length of the bar is half the wave-length and the frequency of the fundamental mode is

$$\nu_0 = \frac{c}{\lambda_0} = \frac{\sqrt{(E/\rho)}}{2l}. \quad \cdot \cdot \cdot \cdot \cdot \cdot (6)$$

An important characteristic of any standing wave is that all particles belonging to a given loop vibrate in the same phase, so that in what follows the system may be replaced by two masses, located at the ends of the bar, each of half the bar's mass. These must be imagined to vibrate in opposite phase and be subject to a restoring force of such magnitude as to reproduce the natural frequency of the bar. We can then apply the much simpler relations for the forced vibrations of a particle (p. 96) to the motion of either half of the bar. If there were no damping at all, the amplitude at resonance would become infinite. In reality, however, there is always at least a small amount of damping present, due both to the resistance of the air at the surface of the bar and to a certain amount of internal friction. The latter is not well understood in terms of atomic theory even for single crystals. It is quite possible that it might vanish altogether for a perfect crystal lattice at absolute zero.

For what follows, the important thing is the phase relation of the driving force to the vibrations of the bar, or of the vibrations of the two particles which replace it in the above sense. The phase relation is given by fig. 7, p. 99. A very graphic representation is obtained by combining both curves of the figure into a single one by the expedient of expressing the amplitude  $a$  of either particle as a function of the phase angle  $\phi$ . For this purpose, eliminate the quantity  $\omega_0^2 - \omega^2$  from equations (33'') and (33'''), p. 98, by using the relation

$$\sin^2 \phi = \frac{\tan^2 \phi}{1 + \tan^2 \phi}. \quad \cdot \cdot \cdot \cdot \cdot \cdot (7)$$

In the notation of p. 98, the result is the polar equation

$$a = \frac{F_0}{\beta \omega} \sin \phi. \quad \cdot \cdot \cdot \cdot \cdot \cdot (8)$$

If the damping constant  $\beta$  is small—as it is for quartz—the region in which the amplitude differs markedly from zero is exceedingly small.



As a result,  $\omega$  may be taken to be constant in this region, and so equation (8) will represent a circle. Since the displacement lags behind the driving force, the circle must be drawn below the polar axis if the driving force  $F$  is drawn along this axis, as shown in fig. 3. This circle then represents the index diagram for the periodic quantities  $F$  and  $u$ , where the latter represents the displacement of one of the particles.

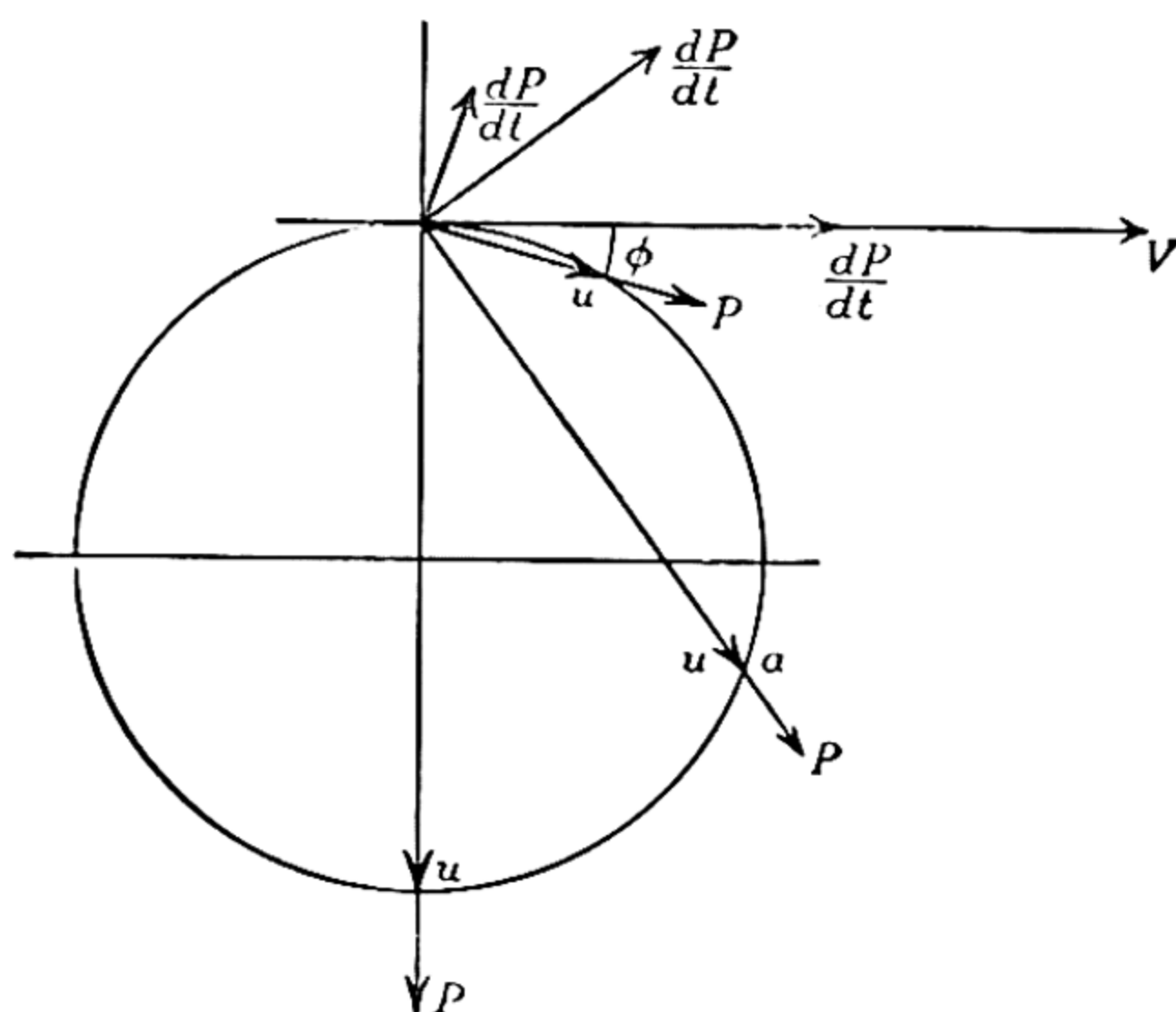


Fig. 3

If an electrical potential of the form  $V = V_0 e^{i\omega t}$  is applied, the driving force is given by

$$F = -\gamma_{11} bc \frac{V}{b}, \quad . . . . . (9)$$

where  $b$  is the width and  $c$  the thickness of the bar (refer to fig. 2). The resonance amplitude becomes

$$a_{\max} = \frac{F_0}{\beta \omega}. \quad . . . . . (10)$$

Since the damping depends on the electrodes as well as on small inhomogeneities of the crystal,  $\beta$  is not to be regarded as a material constant. Rather, the width of the resonance curve is used to characterize the damping of a given crystal.

The polarization  $P$  follows the deformation  $u$  without lag, so that  $P$  has the same direction as  $u$  in the index diagram. The displacement current is given by  $dP/dt = i\omega P$ , so that it leads  $u$  by  $\pi/2$ . If the bar is allowed to pass along the resonance curve by increasing the driving frequency from zero, then  $u$  will have the same direction as  $V$  at low frequencies, but the index of the displacement current will be in the

direction of  $i\omega V$ . Thus the crystal behaves like a pure capacitance, where the current leads the voltage by  $\pi/2$  (p. 318). As the resonance point is approached, the magnitude of  $u$  increases greatly while its index changes to a position  $\pi/2$  behind  $V$  as the resonance point is traversed. At the same time, the direction of the index of the displacement current comes into coincidence with that of  $V$ . The essential result, then, is that at resonance the system behaves like a small ohmic resistance—small because the index of the displacement current becomes very large. The capacitance is effectively short-circuited, i.e. it has become infinitely large. When the resonance point is passed,  $u$  falls even farther behind  $V$ , the displacement current falls behind  $V$ , and the system behaves like an inductance or—what is formally the same thing according to equation (23), p. 318—like a negative capacitance.

In these considerations we have neglected the usually small capacity of the quartz condenser and the displacement current determined by it, since at the critical condition the latter is small compared with the piezoelectric displacement current. Imagine a quartz vibrator connected in parallel with the condenser of an oscillating circuit, the latter being variable so that the circuit can be tuned to any impressed frequency (fig. 4). In the absence of the quartz crystal, the capacitance  $C$  in the relation

$$\omega_0^2 = \frac{1}{LC} \quad . . . . . (11)$$

would have to be made smaller and smaller as the frequency is increased, and the curve connecting  $C$  and  $\omega_0^2$  will be an ordinary hyperbola.

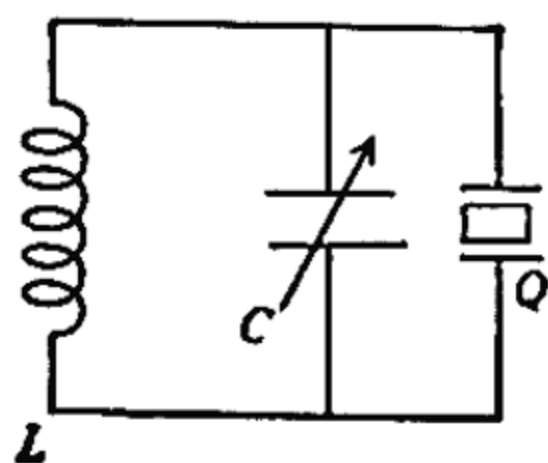


Fig. 4

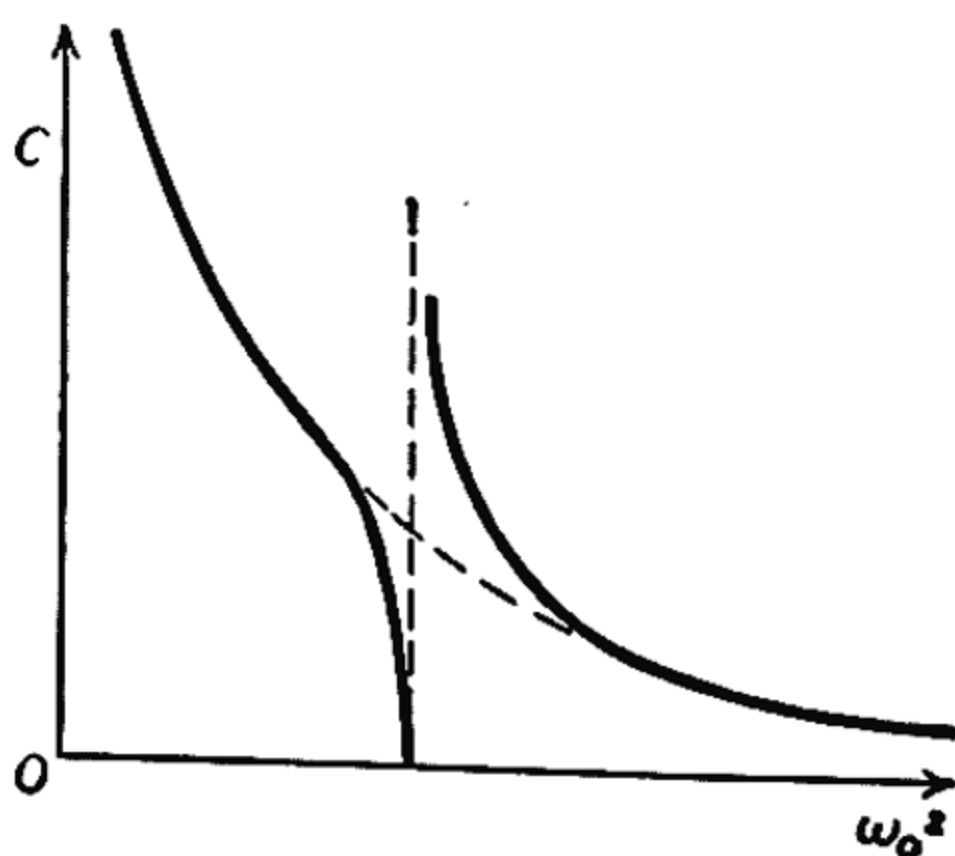


Fig. 5

However, as the resonance point of the crystal is approached, the capacitance of the condenser is increased by that of the quartz, so that  $C$  must be lower than it would be in the absence of the crystal.

At resonance the capacitance of the crystal would become infinite, making it necessary to give  $C$  a negative value large enough to make the sum correspond to  $\omega_0$ .

In practice the ideal conditions are not fulfilled and the capacitance of the crystal, while large, remains finite. The further course of the curve follows from the same considerations (see fig. 5). The stabilizing effect is explained in this way: In order that  $\omega_0^2$  change but little in the neighbourhood of the quartz resonance,  $C$  must change considerably. In other words, a small change in the value of  $C$ —as a result of, say, a change in temperature—will have practically no effect on the natural frequency of the system.



## CHAPTER XLV

### SPACE-CHARGE EFFECTS IN GASEOUS DISCHARGES

#### 1. Characteristic Curve of a Thermionic Tube.

Let us consider an extended cathode, which emits electrons, placed opposite an extended anode, and let us inquire about the relation between current and voltage as qualitatively represented by fig. 1 on p. 433. Because of the indefinite extent of the electrodes, the problem becomes one-dimensional and the results will depend only on a single co-ordinate  $x$ . Further, we may neglect the initial thermal motions of the electrons, whose speeds have approximately a Maxwell distribution (p. 595). With this assumption, the current should be expected to begin only when the second electrode is made positive, while in fact a minute current is found to flow even when there is a small opposing potential.

Under the above assumptions, the values of the potential between the electrodes will be investigated. It will no longer be linear, but will be distorted by the presence of the space charge. We have the following equations at our disposal: The speed of the electrons at any place where the potential with respect to the cathode is  $V$  is given by

$$\frac{1}{2}mv^2 = eV. \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

If  $\rho$  represents the charge density, the relation given on p. 418 yields for the current density \*

$$i = \rho v. \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

Space charge and potential are connected by the relation (p. 267)

$$\Delta V = 4\pi\rho. \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

The reason for the positive sign of the right side of the equation will be understood from the footnote.

The current density  $i$  is constant throughout the space because we assumed the situation to be independent of any but the  $x$  co-ordinate and the current is constant at all points in the circuit. With the help

\* Contrary to general practice, we here take the positive direction of the current to be that of the electron motion.

of (1) and (2) we can eliminate  $\rho$  and  $v$ , and obtain a differential equation for  $V$ :

$$\frac{d^2V}{dx^2} = \frac{4\pi i}{v} = \frac{4\pi i}{\sqrt{\left(\frac{2e}{m}\right) V}} \cdot \cdot \cdot \cdot (4)$$

Multiplying both sides by  $dV/dx$  allows us to integrate at once, and if the potential of the cathode is arbitrarily taken to be zero, the result is

$$\left(\frac{dV}{dx}\right)^2 - \left(\frac{dV}{dx}\right)_0^2 = 8\pi i \sqrt{\left(\frac{2m}{e} V\right)} \cdot \cdot \cdot (5)$$

Under the assumed condition of zero initial speed of the electrons there would be no current if the potential of the cathode were to decrease, for then all the electrons would be driven back. On the other hand, if  $(dV/dx)_0 > 0$ , all the emitted electrons would arrive at the anode and this would constitute the saturation current. Hence the only possibility is to set  $(dV/dx)_0 = 0$  for the region in question. Actually, because of the initial speed of the electrons, the lowest point of the potential curve lies somewhat in front of the cathode, but the above assumption greatly simplifies the computation and does not alter the final result markedly. With this condition assumed, we get

$$\frac{dV}{dx} = (8\pi i)^{\frac{1}{2}} \left(\frac{2m}{e} V\right)^{\frac{1}{4}}, \cdot \cdot \cdot \cdot (6)$$

and by separating the variables we can integrate a second time, obtaining

$$\frac{4}{3} V^{3/4} = (8\pi i)^{1/2} \left(\frac{2m}{e}\right)^{1/4} x \quad \text{or} \quad V \propto x^{4/3} \cdot \cdot \cdot (7)$$

The constant of integration is zero because the potential was assumed to be zero at the cathode. The relation gives the variation of potential between the electrodes, as influenced by the space charge. If  $V_a$  is the anode potential and  $a$  is the separation of the electrodes, this becomes

$$i = \frac{\sqrt{2} \sqrt{(e/m)}}{9\pi a^2} V_a^{3/2} \cdot \cdot \cdot \cdot (8)$$

This well-known "three-halves power law" cannot hold, of course, for arbitrarily large currents. As soon as  $i$  reaches the saturation value there is no further increase, since electrons are being liberated from the cathode at their maximum rate. If the initial electron speed had not been neglected, the  $V_a^{3/2}$  curve would not appear to change sharply over to the saturation line but would have the rounded junction found in practice and pictured on p. 433.





A vibrational differential equation is obtained for the electron velocity as follows: Multiply (10) by  $v$  and add the resulting equation to (12). Then, taking account of the meaning of the differential processes  $\partial/\partial t$  and  $d/dt$  (cf. p. 195), there results

$$\frac{\partial E}{\partial t} + v \frac{\partial E}{\partial x} = \frac{dE}{dt} = 4\pi S + 4\pi e n_0 v. \quad . \quad . \quad . \quad (13)$$

Using (9), we get

$$\frac{d^2 v}{dt^2} + \frac{4\pi e^2 n_0}{m} v = -\frac{4\pi e S}{m}. \quad . \quad . \quad . \quad (14)$$

Now  $S$  is not only constant in space but in time as well. For since the applied field  $E - E_0$  is due to charges, and not to changing magnetic fields,  $\text{curl } \mathbf{E} = 0$ . But the field equations give

$$\text{curl curl } \mathbf{E} = -\frac{4\pi}{c^2} \frac{\partial \mathbf{S}}{\partial t},$$

so that for  $\text{curl } \mathbf{E} = 0$ ,  $\partial \mathbf{S}/\partial t$  must also be zero. The particular integral

$$v = \frac{S}{n_0 e} \quad . \quad . \quad . \quad . \quad . \quad . \quad (15)$$

represents the mean electron current on which is superposed the electron oscillations given by the general integral

$$v = A \sin(\omega t - \alpha), \text{ where } \omega = 2e \sqrt{\left(\frac{\pi n_0}{m}\right)}. \quad . \quad (16)$$

This simple and clear deduction of the oscillations of a plasma is due to Rudolf Seeliger.

## CHAPTER XLVI

### THEORY OF ELASTOMERS

#### 1. Behaviour of High-polymer Chain Molecules.

As a class, the rubber-like solid substances differ greatly from other solids in their mechanical and thermodynamic behaviour. In the phenomenon of expansion, for instance, one must look to quite different molecular processes for an explanation. Apart from the fact that rubber-like substances are more resilient by many orders of magnitude than are other solids, their isobaric coefficient of expansion is negative.\*

Characteristic of all rubber-like materials is the fact that chemically they are high polymers consisting of long chain molecules made up of thousands of identical units. An example is polyacrylic acid, a portion of whose structural formula is shown in fig. 1. The molecule is so long

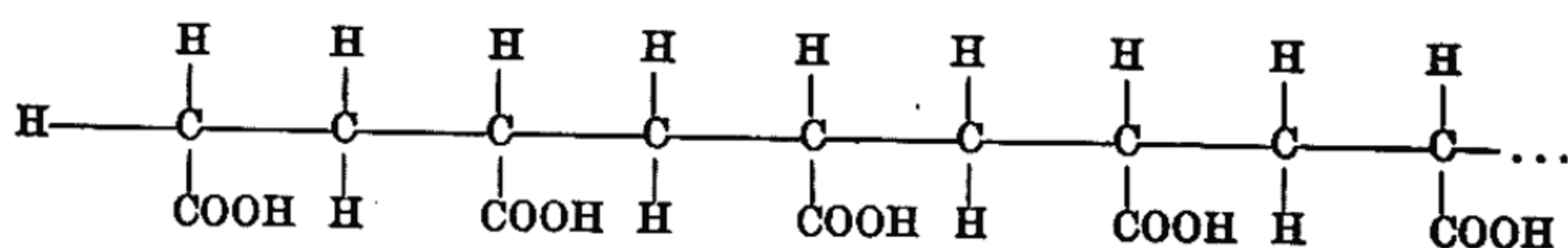


Fig. 1

that probability considerations can be applied even to the parts of a single molecule. Because of the rotational freedom allowed by the simple C—C bond, two adjoining units can, within limits, turn freely with respect to each other. The limiting angle is determined by the structure of the units themselves. After a large number  $n$  of units, the net effect of the many arbitrarily positioned units is such as to make all orientations of the  $n$ th member equally probable. The corresponding length is called the statistical chain element  $A$ . Thus the giant molecule is schematized as a chain of  $N$  such segments whose members may be oriented arbitrarily with respect to one another. The internal energy of the molecule is consequently independent of the form of the chain, which is governed only by the laws of probability. If a perfectly flexible chain is thrown on to a table, it will usually end up in an irregular, coiled configuration; the chance of its assuming the fully-extended position is extremely small. We shall characterize the form solely by

\* This can be demonstrated very simply by hanging a heavy weight by means of a length of rubber tubing. The tube is originally filled with cold water, which is then gradually replaced by hot. Contrary to expectation, the tube shortens.

specifying the straight-line distance between the two ends, which we designate  $h$  (fig. 2). If the beginning of the chain is held fast, a probability calculation gives the following expression for the chance of finding the end of a chain consisting of  $N$  segments somewhere within a spherical shell whose radii are  $h$  and  $h + dh$ :

$$W(h) dh = K e^{-\frac{3h^2}{2NA^2}} h^2 dh. \quad (1)$$

The approximate nature of this expression will be realized when it is seen that values greater than  $NA$ , the fully-extended length, are not excluded, although their probability is given as very small. However, this expression is adequate for our purposes. Suppose a force  $F$  is applied to the end of the chain in the direction  $AB$  (fig. 2). If this

force succeeds in increasing  $h$  by an amount  $dh$ , the work done will be  $F dh$ . According to the first and second laws of thermodynamics, we have for a reversible, isothermal change of state

$$T dS + F dh = dU. \quad (2)$$

However, since the internal energy  $U$  is independent of  $h$ , the right member of the equation is zero and the force is given by

$$F = -T \left( \frac{\partial S}{\partial h} \right)_T. \quad (3)$$

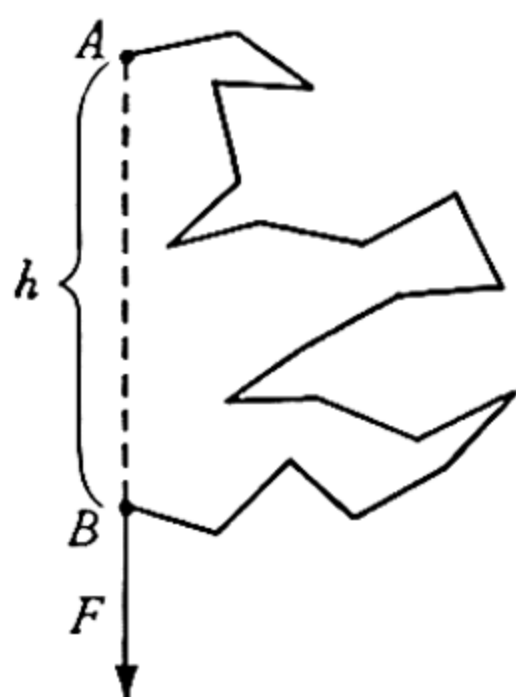


Fig. 2

In order to compute the entropy, we need only apply the Boltzmann equation to the probability formula. The probability that the length of the line  $AB$ , drawn in the direction of  $F$ , shall be between  $h$  and  $h + dh$  is proportional to

$$e^{-\frac{3h^2}{2NA^2}} dh.$$

The factor  $4\pi h^2$  does not enter, since the end of the chain is no longer considered to move on the surface of a sphere, but is restricted to the direction of  $F$ . We then get

$$S = -\frac{3kh^2}{2NA^2} + \text{const.} \quad (4)$$

and

$$F = \frac{3kTh}{NA^2}. \quad (5)$$

Thus the stretching force  $F$  corresponding to a given  $h$  increases with the temperature, or the equilibrium length  $h$  corresponding to a given tension  $F$  decreases with increasing temperature. We saw (p. 575) that



for a crystalline solid the elastic forces are determined by electrical forces between the molecules. For an elastomer, on the other hand, the elasticity is caused only by thermal motion which makes a coiled configuration of the molecule more probable than an extended one.

Now it is obviously not possible to experiment with an individual molecule. A collection of such giant molecules is not a piece of solid rubber but a viscous liquid, since the individual molecules are free to move relative to each other. This is true for the latex and for the sticky raw rubber precipitated from it. Only by the introduction of sulphur atoms (vulcanization) are the individual molecules tied together, permitting us to apply—at least qualitatively—the results obtained for an individual molecule to finite pieces of technically-produced rubber.

## 2. A Model of Muscular Action.

In the contraction of a muscle, nature is able to do something that we have succeeded in doing only indirectly—the immediate conversion of chemical into mechanical energy. Since the amount of free energy thus convertible differs but little from the total energy (cf. p. 553), the efficiency of the process is very high. In the laboratory we can convert the free energy of a galvanic cell into electrical energy which, in principle, could be converted in turn with no loss into mechanical energy. Unfortunately, however, we have not been able to find an electrochemical process for using our greatest source of energy—the oxidation of carbon—for the direct generation of electricity. We are forced to use the roundabout means of heat engines, with their notoriously low efficiencies. In the light of such considerations, the action of animal muscle, which up to now has been quite mysterious, is all the more remarkable. The growing understanding of the nature of giant molecules (proteins belong to this category) offers the possibility of explaining the situation. It has made possible the description of a model for the process whereby chemical energy is converted into mechanical energy.

If the molecular segments mentioned above are now considered to contain electric charges, the internal energy will no longer be independent of  $h$ , and the distribution of lengths will be fundamentally changed. The presence of the electrostatic forces acts to favour the extended configurations in which the charges are as widely separated as possible. Such electrically-charged groups occur when the sodium salt of polyacrylic acid is dissolved in water. The sodium ion  $\text{Na}^+$  dissociates from the  $\text{COONa}$  group to some extent; and instead of the  $\text{COOH}$  groups occurring in the acid, charged  $\text{COO}^-$  groups are present.

A complete computation yields the following results: For uncharged segments, the commonest value of  $h$  is about 0.08 of the fully-extended length of the molecule. With 20 per cent dissociation, on the other hand,

this increases to 0.7 of the maximum length. Suppose the molecule is in a solution whose concentration of  $H^+$  ions (or of  $OH^-$  ions) can be changed. If such a change occurs, the law of mass action will bring about a change in the degree of dissociation of the  $COOH$  group, and the molecule will thus be either shortened or extended—the former if there is an excess of  $H^+$ . In the process, work can be performed by lifting a weight, the energy coming from the chemical energy of neutralization. This model of the muscular process is found to correspond with experience at many points. For example, it correctly predicts that each cubic centimetre of the dry substance should be able to lift about 70,000 times its own weight.

## MATHEMATICAL ADDENDUM

### SOME FUNCTIONS OCCURRING IN CONNEXION WITH THE WAVE EQUATION

#### 1. Generalization of $x!$ for any Value of $x$ .

In the following development we set ourselves the task of finding a function of  $x$  which shall represent the value of a factorial for any kind of  $x$ , including complex values, although this function was originally defined only for  $x$  a positive integer. This function is called  $\Pi(x)$ : it is often designated also as  $\Gamma(x + 1)$ . It is characterized by the relation

$$\Pi(x) = x \Pi(x - 1). \quad . \quad . \quad . \quad . \quad . \quad (1)$$

For positive, real values of  $x$  this condition is satisfied by the integral

$$\Pi(x) = \Gamma(x + 1) = \int_0^\infty t^x e^{-t} dt; \quad . \quad . \quad . \quad . \quad (2)$$

for, integrating by parts,

$$\int_0^\infty t^x e^{-t} dt = -[e^{-t} t^x]_0^\infty + x \int_0^\infty t^{x-1} e^{-t} dt, \quad . \quad . \quad (3)$$

and this is equal to  $x \Pi(x - 1)$ . Now  $\Pi(0) = 1$ , as may be seen by setting  $x$  equal to *one* in equation (1). Hence repeated application of (1) gives, for integral values of the variable,

$$\Pi(x) = \Gamma(x + 1) = 1 \cdot 2 \cdot 3 \cdot 4 \dots x = x!$$

The integral does not converge when  $x$  is negative or when it is complex with a negative real part. For this reason Gauss supplied a generally valid definition of the function  $\Pi(x)$  in the form

$$\Pi(x) = \lim_{n \rightarrow \infty} \frac{n^x n!}{(1 + x)(2 + x)(3 + x) \dots (n + x)}. \quad (4)$$

Obviously, from this definition,  $\Pi(x)$  will be infinite for any integral, negative value of  $x$ . In order to show the identity of both definitions for positive  $x$ , consider the integral

$$J_0(x) = \int_0^1 t^{x+s} (1-t)^{n-s} dt. \quad . \quad . \quad . \quad . \quad (5)$$



When  $n = s$  this becomes

$$J_n(x) = \int_0^1 t^{x+n} dt = \frac{1}{x+n+1}. \quad \dots \quad (6)$$

Integrating (5) by parts,

$$\frac{J_s}{J_{s+1}} = \frac{n-s}{s+x+1}. \quad \dots \quad (7)$$

Repeated application of this recurrence formula yields

$$\begin{aligned} J_0 &= \int_0^1 t^x (1-t)^n dt = \frac{J_0}{J_1} \cdot \frac{J_1}{J_2} \cdot \frac{J_2}{J_3} \dots \frac{J_{n-1}}{J_n} \cdot J_n \\ &= \frac{n!}{(1+x)(2+x)\dots(n+x)} \cdot \frac{1}{n+x+1}, \end{aligned}$$

whence

$$\frac{n^x n!}{(1+x)(2+x)\dots(n+x)} = n^x (n+x+1) \int_0^1 t^x (1-t)^n dt.$$

Setting  $t = u/n$  in the integral,

$$\frac{n^x n!}{(1+x)(2+x)\dots(n+x)} = \frac{n+x+1}{n} \int_0^n \left(1 - \frac{u}{n}\right)^n u^x dx.$$

According to the definition of  $e$  as the limit of  $[1 + (1/n)]^n$  for  $n$  approaching infinity, we have in the limit

$$\lim_{n \rightarrow \infty} \frac{n^x n!}{(1+x)(2+x)\dots(n+x)} = \int_0^\infty u^x e^{-u} du,$$

whereby the agreement of the two definitions for positive values of  $x$  is proved.

It follows from (4) that

$$\begin{aligned} \Pi(-x) \Pi(x-1) &= \lim_{n \rightarrow \infty} \frac{n-1}{x(1-x^2) \left(1 - \frac{x^2}{2^2}\right) \left(1 - \frac{x^2}{3^2}\right) \dots \left(1 - \frac{x^2}{(n-1)^2}\right) (n-x)} \\ &= \lim_{n \rightarrow \infty} \frac{1}{x \left(1 - \frac{x^2}{1^2}\right) \left(1 - \frac{x^2}{2^2}\right) \left(1 - \frac{x^2}{3^2}\right) \dots \left(1 - \frac{x^2}{(n-1)^2}\right)}. \end{aligned}$$

The denominator, however, is merely the familiar infinite product expression for  $\sin \pi x / \pi$ , whence

$$\Pi(-x) \Pi(x-1) = \frac{\pi}{\sin \pi x}. \quad \dots \quad (8)$$

In particular, putting  $x = 1/2$ ,

$$\Pi(-\frac{1}{2}) = \sqrt{\pi}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (9)$$

The expression  $\Psi(x)$  is conventionally used to represent the logarithmic derivative of the  $\Pi$  function. It follows from equation (1) that

$$\Psi(x) = \frac{1}{x} + \Psi(x-1). \quad . \quad . \quad . \quad . \quad . \quad (10)$$

Values of this function may be found in tables. According to the definition (4),

$$\Psi(x) = \frac{\Pi'(x)}{\Pi(x)} = \lim_{n \rightarrow \infty} \left( \log_s x - \frac{1}{1+x} - \frac{1}{2+x} - \dots - \frac{1}{n+x} \right). \quad (11)$$

In particular,

$$\Psi(0) = \lim_{n \rightarrow \infty} \left( \log_e n - \frac{1}{1} - \frac{1}{2} - \frac{1}{3} - \dots - \frac{1}{n} \right) = -C. \quad (12)$$

The negative of the limiting value of the parentheses is called the Euler-Mascheroni constant; its numerical value is  $0.5772 \dots$ . If an integer  $m$  is substituted for  $x$  in (11), the subtracted terms from 1 to  $m$  will be missing in (12) with the result that

$$\Psi(m) = -C + \frac{1}{1} + \frac{1}{2} + \dots + \frac{1}{m}. \quad (13)$$

For negative integral values of  $x$  the function  $\Pi(x)$  goes to infinity as  $1/x$  when  $x$  goes to zero, and so its derivative tends to infinity as  $1/x^2$ . Hence  $\Psi(x)/\Pi(x)$  will be finite, and its value may be deduced as follows: We have

$$\frac{\Psi(x)}{\Pi(x)} = \frac{\Psi(x+1) - \frac{1}{x+1}}{\frac{1}{x+1} \Pi(x+1)} = \frac{[x+1] \left[ \Psi(x+1) - \frac{1}{x+1} \right]}{\Pi(x+1)},$$

from which we get, by repeated application,

$$\frac{\Psi(x)}{\Pi(x)} = \frac{(x+1)(x+2)\dots(x+m) \left[ \Psi(x+m) - \frac{1}{x+1} - \frac{1}{x+2} \dots \frac{1}{x+m} \right]}{\Pi(x+m)}.$$

When  $x = -m$  all terms of the numerator vanish except the one multiplied by  $1/(x + m)$ . Thus, on account of the fact that  $\Pi(0) = 1$ ,

$$\frac{\Psi(-m)}{\Pi(-m)} = (-1)^m \Pi(m-1). \quad (14)$$

## 2. Bessel Functions of the First Kind.

### (a) *The Bessel differential equation*

The wave equation

$$\Delta u = \frac{1}{c^2} \frac{\partial^2 u}{\partial t^2} \quad . \quad . \quad . \quad . \quad . \quad . \quad (15)$$

is transformed in the usual way, by substitution of a periodic solution

$$u = v(x, y, z) e^{i\omega t} \quad . \quad . \quad . \quad . \quad . \quad . \quad (16)$$

into the equation

$$\Delta v + k^2 v = 0, \quad . \quad . \quad . \quad . \quad . \quad . \quad (17)$$

where  $k = \omega/c = 2\pi/\lambda$ . We now investigate the case of cylindrical waves, which are independent of the  $z$  co-ordinate. Then, according to Ex. 22, p. 43, the relation between the two remaining co-ordinates  $\rho$  and  $\phi$  will be

$$\frac{\partial^2 v}{\partial \rho^2} + \frac{1}{\rho} \frac{\partial v}{\partial \rho} + \frac{1}{\rho^2} \frac{\partial^2 v}{\partial \phi^2} + k^2 v = 0. \quad . \quad . \quad . \quad (18)$$

Assume the solution  $v(\rho, \phi)$  to have the form of a product of  $R(\rho)$  and  $\Phi(\phi)$ . In general, only those solutions which are single-valued functions of position are physically valid, so  $\Phi$  must be periodic with a period of  $2\pi$ . Therefore we set

$$\Phi(\phi) = e^{2\pi i p \phi}, \quad . \quad . \quad . \quad . \quad . \quad . \quad (19)$$

where generally, in physical problems,  $p$  is an integer. In some cases, however, non-integral values of  $p$  come into consideration, so that in what follows we shall assume  $p$  to be an arbitrary positive or negative number. Writing for convenience  $y(x)$  in place of  $R(k\rho)$ , we get the equation

$$x^2 \frac{d^2 y}{dx^2} + x \frac{dy}{dx} + (x^2 - p^2)y = 0. \quad . \quad . \quad . \quad (20)$$

This is the *Bessel differential equation*. Apart from certain special values of  $p$ , its solution is not expressible in terms of elementary functions.

### (b) *Solution by power series*

We substitute the power series

$$y = a_1 x^{m_1} + a_2 x^{m_2} + a_3 x^{m_3} + \dots a_r x^{m_r} + \dots \quad . \quad (21)$$

into the differential equation, getting

$$(m_1^2 - p^2)a_1 x^{m_1} + (m_2^2 - p^2)a_2 x^{m_2} + (m_3^2 - p^2)a_3 x^{m_3} + \dots \\ + a_1 x^{m_1+2} + a_2 x^{m_2+2} + a_3 x^{m_3+2} + \dots = 0.$$



In order that the coefficients of each power vanish we must have

$$\left. \begin{aligned} m_2 &= m_1 + 2, \\ m_3 &= m_2 + 2. \end{aligned} \right\} \dots \dots \dots (22)$$

Since the term with the smallest exponent appears only once,

$$\begin{aligned} m_1^2 &= p^2, \\ m_1 &= \pm p. \end{aligned} \dots \dots \dots (23)$$

Let us examine first the solution  $+p$ . From (22),

$$m_2 = p + 2, \quad m_3 = p + 2 \cdot 2, \quad m_r = p + (r - 1)2.$$

We then obtain the coefficients

$$a_2 = -\frac{a_1}{m_2^2 - p^2} = -\frac{a_1}{p^2 + 4p + 4 - p^2} = -\frac{a_1}{2^2(p + 1)},$$

$$a_3 = -\frac{a_2}{m_3^2 - p^2} = -\frac{a_2}{2^2(p + 2) \cdot 2} = \frac{a_1}{2^4 \cdot 2(p + 1)(p + 2)},$$

$$a_4 = -\frac{a_3}{m_4^2 - p^2} = -\frac{a_3}{2^2(p + 3) \cdot 3} = -\frac{a_1}{2^6 \cdot 2 \cdot 3(p + 1)(p + 2)(p + 3)}.$$

From this we generalize to

$$\begin{aligned} a_{k+1} &= -\frac{a_k}{2^2 k(p + k)} \\ &= \frac{(-1)^k a_1}{2^{2k} \Pi(k)(p + 1)(p + 2) \dots (p + k)} = \frac{(-1)^k \Pi(p) a_1}{2^{2k} \Pi(k) \Pi(p + k)}. \end{aligned} \quad (24)$$

The coefficient  $a_1$  is arbitrary, since a multiplicative constant may be inserted into any homogeneous linear differential equation. The normalization of what are called *Bessel functions* (sometimes cylinder functions) of the first kind is attained by setting  $a_1 = 1/2^p \Pi(p)$ . The result is

$$\begin{aligned} J_p &= \frac{x^p}{2^p \Pi(p)} \left[ 1 - \frac{x^2}{2^2(p + 1)} + \frac{x^4}{2^4 \Pi(2)(p + 1)(p + 2)} \right. \\ &\quad \left. - \frac{x^6}{2^6 \Pi(3)(p + 1)(p + 2)(p + 3)} + \dots \right] \\ &= \sum_{k=0}^{\infty} \frac{(-1)^k}{\Pi(k) \Pi(p + k)} \left( \frac{x}{2} \right)^{p+2k} \dots \dots \dots (25) \end{aligned}$$

Comparison with  $\sin x$  and  $\cos x$  shows that the convergence of these

series is excellent. The character of  $J_0$  and of  $J_1$  is to be seen from fig. 1.

In the above derivation there is nothing which restricts the order of the Bessel functions to whole numbers. Even for non-integral values of  $p$  we are able to find at once a second independent solution which supplies the second constant of integration (the general solution of the second-order equation (20) must contain two independent constants). Examining now the value  $m_1 = -p$ , we get an entirely *different* power series. For instance, for  $p = 1/3$ , the exponents  $1/3, 7/3, 13/3 \dots$  occur,

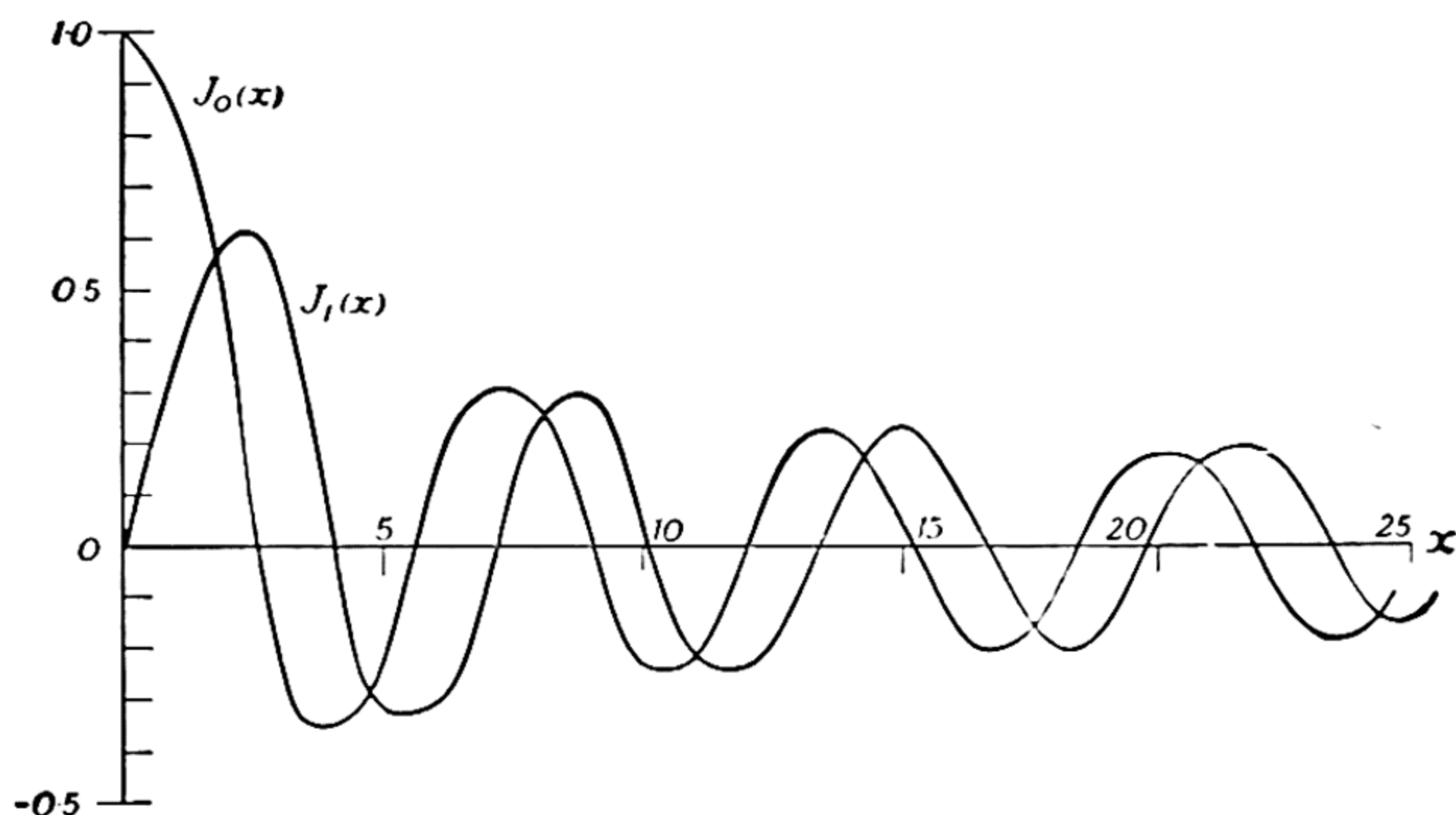


Fig. 1

while for  $p = -1/3$  we get  $-1/3, 5/3, 11/3 \dots$ . Thus for non-integral  $p$ , the general solution of the Bessel equation is given by

$$y = AJ_p + BJ_{-p}. \quad . \quad . \quad . \quad . \quad . \quad (26)$$

The power series behaves quite differently for negative integral values  $-n$  of  $p$ . Since  $\Pi(-n)$  is zero according to equation (4), p. 807, there is no contribution to the sum until  $k = n$ . In order to begin the summation with zero again, we set  $k = l + n$  and obtain the expression

$$J_{-n} = \sum_{l=0}^{\infty} \frac{(-1)^{l+n}}{\Pi(l+n)\Pi(l)} \left(\frac{x}{2}\right)^{n+2l}. \quad . \quad . \quad . \quad (27)$$

But this is identical with  $(-1)^n J_n$ . Hence for integral values  $n$  of the order number  $p$ ,

$$J_{-n} = (-1)^n J_n. \quad . \quad . \quad . \quad . \quad . \quad (28)$$

In the important instance, then, where  $p$  is integral,  $m = -p$  does not supply a second independent solution. This solution must be determined separately (see § 3, below).

(c) *Relations between the Bessel functions and their derivatives*

The Bessel functions have certain similarities with the trigonometric functions; in particular,  $J_0(x)$  is comparable with  $\cos x$ ,  $J_1(x)$  with  $\sin x$ . Like the Bessel functions, the trigonometric functions have infinitely many zeroes. This similarity also exhibits itself in the relations between the derivatives of the functions and the functions themselves, comparable with the relation  $d(\sin x)/dx = \cos x$ .

Let us show, first, that

$$J_p'(x) = \frac{1}{2}[J_{p-1}(x) - J_{p+1}(x)]. \quad . \quad . \quad . \quad (29)$$

By differentiation of the series expression for  $J_p$  we get

$$J_p'(x) = \frac{1}{2} \sum_{k=0}^{\infty} \frac{(p+2k)(-1)^k}{\Pi(k)\Pi(p+k)} \left(\frac{x}{2}\right)^{p+2k-1}.$$

Writing the factor  $p+2k$  in two parts,  $p+k$  and  $k$ , we can obtain two summations. The first yields the series

$$\begin{aligned} \frac{1}{2} \sum_{k=0}^{\infty} \frac{(p+k)(-1)^k}{\Pi(k)\Pi(p+k)} \left(\frac{x}{2}\right)^{p+2k-1} \\ = \sum_{k=0}^{\infty} \frac{(-1)^k}{\Pi(k)\Pi(p+k-1)} \left(\frac{x}{2}\right)^{p+2k-1} = \frac{1}{2} J_{p-1}(x). \end{aligned}$$

The series formed with the second factor will be

$$\frac{1}{2} \sum_{k=0}^{\infty} \frac{k(-1)^k}{\Pi(k)\Pi(p+k)} \left(\frac{x}{2}\right)^{p+2k-1} = \frac{1}{2} \sum_{k=0}^{\infty} \frac{(-1)^k}{\Pi(k-1)\Pi(p+k)} \left(\frac{x}{2}\right)^{p+2k-1}.$$

Since  $1/\Pi(-1) = 0$ , the term corresponding to  $k=0$  will vanish and the denumeration will really begin with  $k=1$ . In order to start counting with zero, we put  $k=l+1$  and obtain

$$-\frac{1}{2} \sum_{l=0}^{\infty} \frac{(-1)^l}{\Pi(l)\Pi(p+l+1)} \left(\frac{x}{2}\right)^{p+2l+1} = -\frac{1}{2} J_{p+1}(x),$$

which proves the relation (29). In particular, since  $J_{-1}(x) = -J_1(x)$ , we get

$$J_0'(x) = J_1(x). \quad . \quad . \quad . \quad . \quad (29a)$$

In the same way, we find that

$$J_p'(x) = -\frac{p}{x} J_p(x) + J_{p-1}(x). \quad . \quad . \quad . \quad . \quad (30)$$

From (29) and (30) it follows that

$$\frac{2p}{x} J_p(x) = J_{p-1}(x) + J_{p+1}(x). \quad . \quad . \quad . \quad . \quad (31)$$



A variety of further relations may be deduced from the recurrence formulæ (29) to (31), but none of these will be given here. An important integral relation, however, will be derived. Multiplication of (30) by  $x^p$  yields

$$x^p J_p'(x) + px^{p-1} J_p(x) = x^p J_{p-1}(x).$$

The left side of the equation is the exact derivative of  $x^p J_p(x)$ , so that the result of integration is

$$\int x^p J_{p-1}(x) dx = x^p J_p(x). \quad . \quad . \quad . \quad (32)$$

(d) *Representation of Bessel functions of integral order by definite integrals*

It will be shown that for integer values  $n$  of  $p$ , the fundamental differential equation (20), p. 810, is satisfied by the function

$$J_n(x) = \frac{1}{\pi} \int_0^\pi \cos(x \sin \phi - n\phi) d\phi. \quad . \quad . \quad . \quad (33)$$

Differentiating,

$$\begin{aligned} \frac{dJ_n(x)}{dx} &= -\frac{1}{\pi} \int_0^\pi \sin(x \sin \phi - n\phi) \sin \phi d\phi; \\ \frac{d^2 J_n(x)}{dx^2} &= -\frac{1}{\pi} \int_0^\pi \cos(x \sin \phi - n\phi) \sin^2 \phi d\phi. \end{aligned}$$

Substituting in the differential equation, we must have

$$\begin{aligned} \int_0^\pi \cos(x \sin \phi - n\phi) \sin^2 \phi d\phi + \frac{1}{x} \int_0^\pi \sin(x \sin \phi - n\phi) \sin \phi d\phi \\ + \int_0^\pi \left( \frac{n^2}{x^2} - 1 \right) \cos(x \sin \phi - n\phi) d\phi = 0. \end{aligned}$$

To prove the validity of this equation, transform the second integral by integrating by parts, obtaining

$$\begin{aligned} \int_0^\pi \sin(x \sin \phi - n\phi) \sin \phi d\phi &= - \left[ \sin(x \sin \phi - n\phi) \cos \phi \right]_0^\pi \\ &+ \int_0^\pi \cos \phi \cos(x \sin \phi - n\phi) (x \cos \phi - n) d\phi. \end{aligned}$$

Combining with other integrals, we get simply

$$- \frac{n}{x^2} \int_0^\pi \cos(x \sin \phi - n\phi) (x \cos \phi - n) d\phi = - \frac{n}{x^2} \left[ \sin(x \sin \phi - n\phi) \right]_0^\pi$$

which turns out to be zero, thus proving that the function (33) satisfies the Bessel equation.

(e) Orthogonality of the Bessel functions *My*

The computation of the coefficients of Fourier series is made possible by the vanishing of certain integrals of trigonometric functions. The Bessel functions have similar properties, so that series developments can be set up in terms of Bessel functions instead of in terms of trigonometric functions. This may be shown by starting with the following differential equations:

$$x^2 y_1'' + x y_1' + (a^2 x^2 - p^2) y_1 = 0,$$

$$x^2 y_2'' + x y_2' + (\beta^2 x^2 - p^2) y_2 = 0.$$

By making the substitutions  $z = ax$ ,  $z = \beta x$ , it is easy to see that the solution of the first equation is  $y_1 = J_p(ax)$ , and that of the second is  $y_2 = J_p(\beta x)$ . Multiply the first equation by  $y_2/x$ , the second by  $y_1/x$ , and subtract:

$$x(y_1'' y_2 - y_2'' y_1) + (y_1' y_2 - y_2' y_1) + (a^2 - \beta^2) x y_1 y_2 = 0.$$

Multiply by  $dx$  and integrate from 0 to 1. Then, since the first two parentheses taken together are exactly equal to the differential of  $x(y_1' y_2 - y_2' y_1)$ , we have

$$(a^2 - \beta^2) \int_0^1 x y_1 y_2 dx = - \left[ x(y_1' y_2 - y_2' y_1) \right]_0^1.$$

Substituting the Bessel functions in place of  $y$  and remembering that  $y_1' = a J_p'(ax)$ , there results

$$\begin{aligned} (a^2 - \beta^2) \int_0^1 x J_p(ax) J_p(\beta x) dx \\ = \left[ x \{ \beta J_p(ax) J_p'(\beta x) - a J_p(\beta x) J_p'(ax) \} \right]_0^1. \end{aligned}$$

If we exclude functions of fractional, negative orders,  $p < -1$ , which become infinite of too high an order when  $x = 0$ , the expression in brackets will vanish at the lower limit and we get

$$\int_0^1 x J_p(ax) J_p(\beta x) dx = \frac{\beta J_p(a) J_p'(\beta) - a J_p(\beta) J_p'(a)}{a^2 - \beta^2}.$$

If, now,  $a$  and  $\beta$  are two different zeroes of the Bessel function, we have the first orthogonality equation,

$$\int_0^1 x J_p(ax) J_p(\beta x) dx = 0.$$

The expression becomes indeterminate when  $\alpha = \beta$ . Assuming  $\beta$  to differ from  $\alpha$  by a small quantity which is allowed to approach zero, the value for  $\alpha = \beta$  may be obtained by differentiating numerator and denominator with respect to  $\beta$ , and then setting  $\beta = \alpha$ . The result is

$$\int_0^1 x J_p(\alpha x) J_p(\alpha x) dx = \frac{\alpha J_p'(\alpha) J_p'(\alpha) - J_p(\alpha) J_p'(\alpha) - \alpha J_p(\alpha) J_p''(\alpha)}{2\alpha}.$$

Using the fact that  $J_p(\alpha) = 0$ , we get the second orthogonality equation

$$\int_0^1 x J_p^2(\alpha x) dx = \frac{1}{2} J_p'^2(\alpha). \quad . \quad . \quad . \quad (35)$$

### 3. Bessel Functions of the Second and Third Kinds. Asymptotic Values of Bessel Functions.

#### (a) Expansion of Bessel functions of the second kind in series

When integers are substituted for  $p$  in equation (25) no new function is obtained. To find this result, we pass to the limit by letting an order number which differs slightly from a whole number approach the integer value. Consider the expression

$$N_p(x) = \frac{J_p(x) \cos p\pi - J_{-p}(x)}{\sin p\pi}, \quad . \quad . \quad . \quad (36)$$

which certainly represents a solution of the Bessel equation. When  $p$  is an integer, both numerator and denominator vanish, and the limiting value is obtained by differentiating numerator and denominator. The result is

$$\begin{aligned} N_n(x) &= \frac{-\pi \sin p\pi J_p(x) + \cos p\pi \frac{\partial J_p(x)}{\partial p} - \frac{\partial J_{-p}(x)}{\partial p}}{\pi \cos p\pi} \Big|_{p=n} \\ &= \frac{(-1)^n}{\pi} \left[ (-1)^n \frac{\partial J_p(x)}{\partial p} - \frac{\partial J_{-p}(x)}{\partial p} \right]_{p=n}. \quad . \quad . \quad (37) \end{aligned}$$

Using the notation of § 1, we get, further,

$$\begin{aligned} \left[ \frac{\partial J_p(x)}{\partial p} \right]_{p=n} &= \sum_{k=0}^{\infty} \frac{(-1)^k}{\Pi(k) \Pi(n+k)} \left( \frac{x}{2} \right)^{n+2k} \left[ \log_e \frac{x}{2} - \Psi(n+k) \right], \\ \left[ \frac{\partial J_{-p}(x)}{\partial p} \right]_{p=n} &= \sum_{k=0}^{\infty} \frac{(-1)^k}{\Pi(k) \Pi(-n+k)} \left( \frac{x}{2} \right)^{-n+2k} \left[ -\log_e \frac{x}{2} + \Psi(-n+k) \right]. \end{aligned}$$



Multiplying by  $(-1)^n$  and subtracting, we have, by (14),

$$\pi N_n(x) = 2J_n(x) \log_e \frac{x}{2} - \sum_{k=0}^{\infty} \frac{(-1)^k \Psi(n+k)}{\Pi(k) \Pi(n+k)} \left(\frac{x}{2}\right)^{n+2k} - \sum_{k=0}^{\infty} \frac{(-1)^{k+n} \Psi(-n+k)}{\Pi(k) \Pi(-n+k)} \left(\frac{x}{2}\right)^{-n+2k}. \quad (38)$$

As long as  $k \geq n$ , the same powers of  $x$  occur in the first summation in connexion with  $n+k$  as do in the second one in connexion with  $k$ . The coefficients are  $\Psi(k)/\Pi(k+n)\Pi(k)$ , and if these are combined with the first summation, all that remains of the second one is the part from  $k=0$  to  $k=n-1$ . Here the coefficients are [cf. equation (14) of § 1]:

$$\frac{\Psi(-n+k)}{\Pi(k) \Pi(-n+k)} = (-1)^{n-k} \frac{\Pi(n-k-1)}{\Pi(k)}.$$

Finally,

$$N_n(x) = \frac{2}{\pi} J_n \log_e \frac{x}{2} - \frac{1}{\pi} \sum_{k=0}^{\infty} (-1)^k \frac{\Psi(k) + \Psi(n+k)}{\Pi(k) \Pi(n+k)} \left(\frac{x}{2}\right)^{n+2k} - \frac{1}{\pi} \sum_{k=0}^{n-1} \frac{\Pi(n-k-1)}{\Pi(k)} \left(\frac{x}{2}\right)^{-n+2k}. \quad (39)$$

It follows from this development that the Bessel functions of the second kind (also called Neumann functions) become infinite as  $x^{-n}$  when  $x$  approaches zero. For  $N_0$  the sum involving negative powers drops out and the first term makes the entire function approach an infinite value logarithmically (see fig. 2).

### (b) Bessel functions of the third kind (Hankel functions)

Bessel functions of the third kind are defined by

$$\left. \begin{aligned} H_p^{(1)}(x) &= J_p(x) + iN_p(x), \\ H_p^{(2)}(x) &= J_p(x) - iN_p(x). \end{aligned} \right\} \quad \dots \dots \dots (40)$$

As will be shown immediately, these functions are related to  $J_p$  and  $N_p$  in the same way that  $e^{iz}$  is related to  $\sin x$  and  $\cos x$ .

### (c) Asymptotic forms of the Bessel functions

The differential equation (20) may be reduced to a simpler form by setting

$$y = x^m z \quad \dots \dots \dots (41)$$

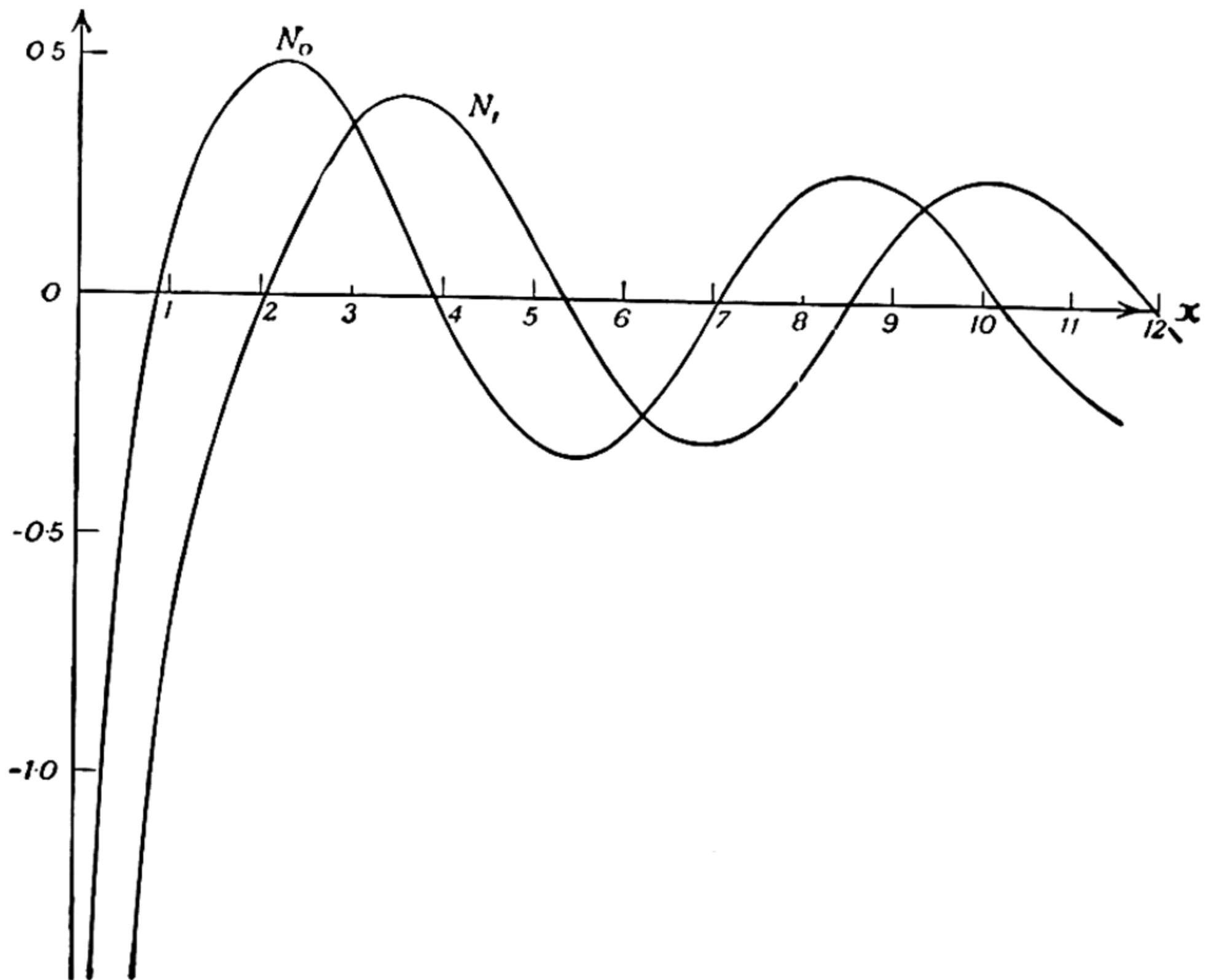


Fig. 2

and determining the exponent  $m$  in such way as to make the term containing the first derivative of  $z$  vanish. As may be readily verified, this occurs when  $m = -1/2$ , leaving the following equation for  $z$ :

$$\frac{d^2z}{dx^2} + z \left( 1 - \frac{p^2 - \frac{1}{4}}{x^2} \right) = 0. \quad . \quad . \quad . \quad (42)$$

Obviously when  $p = \pm 1/2$  this becomes the differential equation of simple harmonic vibration and  $J_{1/2}$  as well as  $J_{-1/2}$  must be of the form  $(A \sin x + B \cos x)/\sqrt{x}$ . Comparison of the series expressions yields the correct values of the constants:

$$\left. \begin{aligned} J_{1/2}(x) &= \sqrt{\left(\frac{2}{\pi x}\right)} \cdot \sin x, \\ J_{-1/2}(x) &= \sqrt{\left(\frac{2}{\pi x}\right)} \cdot \cos x. \end{aligned} \right\} \quad . \quad . \quad . \quad (43)$$

However, for any value of  $p$  we again get the harmonic equation if  $x$  is made large enough. The asymptotic values of the Bessel functions become, after determining the constants,

$$\left. \begin{aligned} H_p^{(1)}(x) &\doteq \sqrt{\left(\frac{2}{\pi x}\right)} \cdot e^{ix - i\pi\left(\frac{2p+1}{4}\right)} \\ H_p^{(2)}(x) &\doteq \sqrt{\left(\frac{2}{\pi x}\right)} \cdot e^{-ix + i\pi\left(\frac{2p+1}{4}\right)} \end{aligned} \right\} \text{ for } x \text{ large.}$$

That is,

$$J_p(x) \doteq \sqrt{\left(\frac{2}{\pi x}\right)} \cdot \cos \left[ x - \pi \left( \frac{2p+1}{4} \right) \right] \text{ for } x \text{ large.}$$

This leads to a way of representing cylindrical waves. If, in accordance with equation (16), p. 810, we set

$$u = J(k\rho)e^{i\omega t},$$

all particles will vibrate in the same phase and we have a standing wave. A progressive wave will be characterized by the use of  $\omega(t - x/v)$  as the argument of the periodic function. As the asymptotic form shows, this is accomplished by the Hankel functions  $H_p^{(2)}$  in conjunction with  $e^{i\omega t}$ .

*Ex. 135.* Represent  $J_{3/2}$  by means of elementary functions.

#### 4. Spherical Harmonics.

##### (a) *Differential equation of spherical harmonics*

The Laplacian potential equation  $\Delta u = 0$  leads at once to another class of functions. We seek such solutions as are homogeneous and of degree  $n$  in the co-ordinates  $x, y, z$ . These solutions, expressed in three-dimensional polar co-ordinates, are of the form

$$u = r^n F_n(\theta, \phi). \quad . \quad . \quad . \quad . \quad . \quad (44)$$

We write the potential equation in polar form and substitute (44). Then, according to p. 268, the equation for  $F_n$  becomes

$$\frac{1}{\sin \theta} \frac{\partial}{\partial \theta} \left( \sin \theta \frac{\partial F_n}{\partial \theta} \right) + \frac{1}{\sin^2 \theta} \frac{\partial^2 F_n}{\partial \phi^2} + n(n+1)F_n = 0. \quad (45)$$

Inasmuch as the equation remains unchanged if  $-(n+1)$  replaces  $n$ ,  $r^{-(n+1)}F_n$  is also a solution if  $r^n F_n$  is a solution.

Assuming independence of  $\phi$ , a simpler equation results by setting  $\cos \theta = x$  and  $F_n = y$ :

$$\frac{d}{dx} \left[ (1-x^2) \frac{dy}{dx} \right] + n(n+1)y = 0. \quad . \quad . \quad . \quad (46)$$



(b) *Legendre polynomials*

As in the case of Bessel functions, we can find a power-series solution for the differential equation. It is found that for one of the two solutions the power series terminates, i.e. the solution is given by a polynomial. However, a more vivid way of arriving at our goal is the following: Consider a unit electrostatic charge located on the  $z$  axis at unit distance from the origin. At a point  $P$  whose distance from the charge is  $r'$ , the electrostatic potential will be

$$u = \frac{1}{r'} = (1 + r^2 - 2r \cos \theta)^{-1/2} = [1 + r(r - 2x)]^{-1/2}. \quad (47)$$

If  $r$  is small, the radical may be expanded by the binomial theorem, yielding a series in powers of  $r$ . Since each term is a solution of the Laplace equation, the coefficients will be the *Legendre polynomials* (or *surface zonal harmonics*)  $P_n$  that we seek. Denoting the largest integer which is less than or equal to  $n/2$  by  $[n/2]$ , the general expression for these functions may be written

$$P_n(x) = \sum_{k=0}^{[n/2]} (-1)^k \frac{1.3.5 \dots (2n - 2k - 1)}{2^k k! (n - 2k)!} x^{n-2k}. \quad (48)$$

The polynomials for  $n = 0$  to  $n = 5$ , inclusive, were given on p. 699. Since  $x = \cos \theta$ , we obtain  $P_n$  in powers of  $\cos \theta$ . Using the well-known formulæ, these functions may be expressed in terms of multiples of the angle. For example, for  $P_2$ ,

$$P_2(x) = \frac{1}{2}(3x^2 - 1) = \frac{1}{4}(3 \cos 2\theta + 1).$$

(c) *Relations between spherical harmonics of various orders*

The development of (47), arranged in ascending powers of  $r$ , has the form

$$(1 - 2rx + r^2)^{-1/2} = P_0(x) + rP_1(x) + r^2P_2(x) + \dots r^n P_n(x) + r^{n+1}(P_{n+1}(x) + \dots)$$

Differentiating throughout with respect to  $r$ ,

$$(1 - 2rx + r^2)^{-3/2}(x - r) = P_1(x) + 2rP_2(x) + \dots nr^{n-1}P_n(x) + (n+1)r^n P_{n+1}(x) + \dots$$

or

$$(1 - 2rx + r^2)^{-1/2}(x - r) = (P_1(x) + 2rP_2(x) + \dots)(1 - 2rx + r^2).$$

Developing the left side once more according to powers of  $r$ ,

$$(x - r)(P_0(x) + rP_1(x) + \dots + r^n P_n(x) + r^{n+1} P_{n+1}(x) + \dots) \\ = (P_1(x) + 2rP_2(x) + \dots)(1 - 2rx + r^2).$$

Comparison of the same powers of  $r$  yields the relation

$$(2n + 1)xP_n(x) = (n + 1)P_{n+1}(x) + nP_{n-1}(x). \quad (49)$$

In the same way we obtain, by differentiating with respect to  $x$ ,

$$(2n + 1)P_n(x) = P_{n+1}'(x) - P_{n-1}'(x). \quad (50)$$

### (d) The associated harmonics

We now take up the consideration of a *single-valued* dependence of  $F_n$  on the azimuth  $\phi$ , setting

$$F_n(\theta, \phi) = e^{im\phi} f_n(\theta). \quad (51)$$

We then obtain as the differential equation for  $f_n$

$$\frac{1}{\sin \theta} \frac{d}{d\theta} \left[ \sin \theta \frac{df_n}{d\theta} \right] + \left[ n(n + 1) - \frac{m^2}{\sin^2 \theta} \right] f_n = 0. \quad (52)$$

Taking  $f_n(\cos \theta) = y(x)$ , this gives

$$(1 - x^2) \frac{d^2 y}{dx^2} - 2x \frac{dy}{dx} + \left[ n(n + 1) - \frac{m^2}{1 - x^2} \right] y = 0. \quad (53)$$

In order to simplify this relation, we set  $y = (1 - x^2)^p u_m$  and determine  $p$  in such a way as to make the coefficient of  $u_m$  a constant. The result is

$$p = \frac{1}{2}m, \quad y = (1 - x^2)^{m/2} U_m, \quad (54)$$

and the differential equation for  $u_m$  becomes

$$(1 - x^2) \frac{d^2 u_m}{dx^2} - 2(m + 1)x \frac{du_m}{dx} + [n(n + 1) - m(m + 1)] u_m = 0. \quad (55)$$

Differentiating,

$$(1 - x^2) \frac{d^3 u_m}{dx^3} - 2(m + 2)x \frac{d^2 u_m}{dx^2} + [n(n + 1) - (m + 1)(m + 2)] \frac{du_m}{dx} = 0. \quad (56)$$

Thus if  $u_m$  satisfies (55) for  $m$ , then  $du_m/dx$  satisfies (55) for  $m + 1$ .

For  $m = 0$ ,  $u_0$  becomes identical with  $P_n$ , and so  $d^m P_n(x)/dx^m$  satisfies (55) for  $m = m$ , and the solution of (53) for  $m = m$  is

$$P_n^m(x) = (1 - x^2)^{m/2} \frac{d^m P_n(x)}{dx^m}. \quad (57)$$

These functions are called the associated *Legendre polynomials* or *associated harmonics*. Since the  $n$ th derivative of a polynomial of  $n$ th degree is a constant, we have

$$P_n^n(x) = C(1 - x^2)^{m/2} = C \sin^n \theta$$

and

$$P_n^m(x) = 0 \text{ for } m > n.$$

If (49) is differentiated  $m$  times, the result is

$$\begin{aligned} (2n+1)x \frac{d^m P_n(x)}{dx^m} + m(2n+1) \frac{d^{m-1} P_n(x)}{dx^{m-1}} \\ = (n+1) \frac{d^m P_{n+1}(x)}{dx^m} + n \frac{d^m P_{n-1}(x)}{dx^m}. \end{aligned}$$

Now if (50) is differentiated  $(m-1)$  times, we get

$$(2n+1) \frac{d^{m-1} P_n(x)}{dx^{m-1}} = \frac{d^m P_{n+1}(x)}{dx^m} - \frac{d^m P_{n-1}(x)}{dx^m}.$$

Multiplying the last equation by  $m$  and subtracting the result from the preceding equation, we get, after multiplying by  $(1-x^2)^{m/2}$ ,

$$(2n+1)xP_n^m(x) = (n-m+1)P_{n+1}^m(x) + (n+m)P_{n-1}^m(x). \quad (58)$$

### (e) Orthogonality of the Legendre polynomials

Consider two Legendre polynomials having the same upper index but different lower indices  $n$  and  $n'$  and satisfying the differential equations

$$\begin{aligned} \frac{d}{dx} \left[ (1-x^2) \frac{dP_n^m(x)}{dx} \right] + \left[ n(n+1) - \frac{m^2}{1-x^2} \right] P_n^m(x) &= 0, \\ \frac{d}{dx} \left[ (1-x^2) \frac{dP_{n'}^m(x)}{dx} \right] + \left[ n'(n'+1) - \frac{m^2}{1-x^2} \right] P_{n'}^m(x) &= 0. \end{aligned}$$

Multiply the first of these by  $P_{n'}^m$ , the second by  $P_n^m$ , and subtract the second from the first. Integrating the result between the limits  $-1$  and  $+1$ , we get

$$\begin{aligned} [n(n+1) - n'(n'+1)] \int_{-1}^{+1} P_n^m(x) P_{n'}^m(x) dx = \\ \int_{-1}^{+1} P_{n'}^m(x) \frac{d}{dx} \left[ (1-x^2) \frac{dP_n^m(x)}{dx} \right] dx - \int_{-1}^{+1} P_n^m(x) \frac{d}{dx} \left[ (1-x^2) \frac{dP_{n'}^m(x)}{dx} \right] dx. \end{aligned} \quad (59)$$



Integrating by parts, the first integral on the right becomes

$$\int_{-1}^{+1} P_{n'}^m(x) \frac{d}{dx} \left[ (1-x^2) \frac{dP_n^m(x)}{dx} \right] dx = \left[ (1-x^2) P_{n'}^m(x) \frac{dP_n^m(x)}{dx} \right]_{-1}^{+1} - \int_{-1}^{+1} (1-x^2) \frac{dP_n^m(x)}{dx} \frac{dP_{n'}^m(x)}{dx} dx.$$

The bracketed expression vanishes at both limits, while the integral expression cancels the corresponding one that is obtained in a similar way from the second integral in (59). Hence

$$\int_{-1}^{+1} P_{n'}^m(x) P_n^m(x) dx = 0 \text{ for } n' \neq n. \quad \cdot \cdot \quad (60)$$

The computation of  $\int_{-1}^{+1} [P_n^m(x)]^2 dx$  is somewhat more troublesome, and so we content ourselves with quoting the result:

$$\int_{-1}^{+1} [P_n^m(x)]^2 dx = \frac{2}{2n+1} \frac{(n+m)!}{(n-m)!}. \quad \cdot \cdot \quad (61)$$

As in wave mechanics, the proper functions—mostly Legendre polynomials—are assumed to be “normalized to unity”, the values given in tables must be multiplied by the reciprocal of the square root of the preceding expression if numerical values are of concern in the problem.

*(f) Application to the wave-mechanical calculation of the selection rules for the magnetic and azimuthal quantum numbers*

In the older quantum theory, the selection rules were deduced by applying the correspondence principle to the atom model. In wave mechanics, these rules are consequences of the orthogonality properties of spherical harmonics. Imagine a unique direction chosen in an arbitrarily weak magnetic field, and let this direction be the  $z$  axis. Neglecting normalization, two proper functions of hydrogen may be written in the form (p. 702)

$$\psi_i = R_i(r) P_{l_i}^{m_i}(\cos\theta) \cos m_i \phi \quad (i = 1 \text{ or } 2).$$

According to p. 705, the dipole moment governing the radiation accompanying the transition from 1 to 2 will be

$$\mathbf{p} = e \int \mathbf{r} \psi_1 \psi_2 d\tau,$$

where the integration is to be extended over the entire space. Since  $z = r \cos\theta$ , the  $z$  component of this moment may be written out in

the form of a product of three integrals:

$$p_z = e \int_0^\infty r R_1(r) R_2(r) r^2 dr \\ \times \int_0^\pi P_{l_1}^{m_1}(\cos \theta) P_{l_2}^{m_2}(\cos \theta) \cos \theta \sin \theta d\theta \\ \times \int_0^{2\pi} \cos m_1 \phi \cos m_2 \phi d\phi. \quad . \quad . \quad . \quad (62)$$

The last integral vanishes, except when  $m_1 = m_2$ . This is the selection rule applying to the magnetic quantum number for light vibrating parallel to the field direction. For  $m_1 = m_2 = m$ , the second integral, written in terms of the variable  $x$ , becomes

$$J_2 = \int_{-1}^{+1} x P_{l_1}^m(x) P_{l_2}^m(x) dx. \quad . \quad . \quad . \quad (63)$$

According to equation (58), this becomes

$$J_2 = \int_{-1}^{+1} \frac{l_1 - m + 1}{2l_1 + 1} P_{l_1}^m(x) P_{l_2}^m(x) dx + \int_{-1}^{+1} \frac{l_1 + m}{2l_1 + 1} P_{l_1-1}^m(x) P_{l_2}^m(x) dx. \quad (64)$$

The first integral is zero except when  $l_2 = l_1 + 1$ ; the second is zero except when  $l_2 = l_1 - 1$ . This is nothing more nor less than the selection rule  $l \rightarrow l \pm 1$ .

### (g) Legendre functions of the second kind

For the sake of completeness it should be added that the Legendre polynomials, containing as they do only *one* multiplicative constant, represent only one of the pair of solutions of the differential equation. A second solution, for  $m = 0$  and  $n = 0$ , is

$$Q_0(x) = \tanh^{-1} x, \quad . \quad . \quad . \quad . \quad (65)$$

and for  $m = 0$  and  $n = 1$ ,

$$Q_1(x) = x \tanh^{-1} x - 1. \quad . \quad . \quad . \quad . \quad (66)$$

The succeeding Legendre functions of the second kind are obtained from the following recurrence formula, corresponding to equation (49), p. 821:

$$(n+1)Q_{n+1}(x) = (2n+1)xQ_n(x) - nQ_{n-1}(x). \quad . \quad (67)$$

In the same way, the associated Legendre functions of the second kind are obtained from

$$Q_n^m(x) = (1-x^2)^{m/2} \frac{d^m Q_n(x)}{dx^m}. \quad . \quad . \quad . \quad (68)$$

# APPENDIX

## SOLUTIONS OF THE EXERCISES

The more difficult exercises are marked with an asterisk. In such instances it is intended that the solution should be consulted as an aid in solving the problem.

*Ex. 1, p. 10.*  $\mathbf{A} + \mathbf{B} + \mathbf{C} = 0$ .

*Ex. 2, p. 10.* (a)  $\mathbf{B} = \alpha\mathbf{A}$ ; (b)  $\mathbf{C} = \alpha\mathbf{A} + \beta\mathbf{B}$ , where  $\alpha$  and  $\beta$  are scalars.

*Ex. 3, p. 12.* This is the trigonometric Law of Cosines. The alteration in sign of the term  $2AB \cos(\mathbf{A}, \mathbf{B})$  is due to the fact that here  $\cos(\mathbf{A}, \mathbf{B})$  refers to the exterior angle of the triangle.

*Ex. 4, p. 12.*  $(\mathbf{A} + \mathbf{B})$  and  $(\mathbf{A} - \mathbf{B})$  are the diagonals of the parallelogram formed by  $\mathbf{A}$  and  $\mathbf{B}$ . In the particular instance where  $A^2 = B^2$  there results the theorem that the diagonals of a rhombus are mutually perpendicular.

*Ex. 5, p. 12.*  $\cos \theta = s_1 s_2 = \cos \alpha_1 \cos \alpha_2 + \cos \beta_1 \cos \beta_2 + \cos \gamma_1 \cos \gamma_2$ .

*Ex. 6, p. 15.* Project the vectors  $\mathbf{B}$ ,  $\mathbf{C}$  and  $(\mathbf{B} + \mathbf{C})$  on the plane normal to  $\mathbf{A}$ , in which the product vectors  $[\mathbf{AB}]$ , &c., lie. The product vectors are obtained from the projections by rotation through  $\pi/2$  and multiplication by  $|\mathbf{A}|$ , since the magnitude of the projection of  $\mathbf{B}$  is  $|\mathbf{B}| \sin(\mathbf{AB})$ . Since the projection of the sum of vectors is equal to the sum of the projections, and since the altered parallelogram of the projections is similar to the original form, (16) is proved.

*Ex. 7, p. 15.*  $A^2 B^2 \sin^2 \theta + A^2 B^2 \cos^2 \theta = A^2 B^2$ .

*Ex. 8, p. 15:*

(a) Since the projection upon  $\mathbf{n}$  of a radius vector drawn to any point of the plane must be equal to  $p$ , the equation of the plane must be

$$\mathbf{r}\mathbf{n} = p \quad \text{or} \quad x \cos \alpha + y \cos \beta + z \cos \gamma - p = 0.$$

(b)  $p_0 = (\mathbf{r}_0 - \mathbf{r})\mathbf{n} = \mathbf{r}_0\mathbf{n} - p$ .

(c) The vectors  $(\mathbf{r} - \mathbf{r}_1)$ ,  $(\mathbf{r}_2 - \mathbf{r}_1)$  and  $(\mathbf{r}_3 - \mathbf{r}_1)$  are to be coplanar, i.e. the parallelepiped formed by them must have zero volume, whence the equation of the plane is

$$(\mathbf{r} - \mathbf{r}_1)[(\mathbf{r}_2 - \mathbf{r}_1)(\mathbf{r}_3 - \mathbf{r}_1)] = 0,$$

or, in terms of the co-ordinates, we get the equation of the plane:

$$\begin{vmatrix} x - x_1 & y - y_1 & z - z_1 \\ x_2 - x_1 & y_2 - y_1 & z_2 - z_1 \\ x_3 - x_1 & y_3 - y_1 & z_3 - z_1 \end{vmatrix} = 0.$$

*Ex. 9, p. 17.* If we set  $[\mathbf{AB}] = \mathbf{E}$  and apply (24) we obtain

$$\mathbf{C} \cdot \mathbf{ED} - \mathbf{D} \cdot \mathbf{EC} = \mathbf{C} \cdot \mathbf{ABD} - \mathbf{D} \cdot \mathbf{ABC},$$

i.e. a vector in the plane of  $\mathbf{C}$  and  $\mathbf{D}$ . Similarly, putting  $[\mathbf{CD}] = \mathbf{F}$  yields  $\mathbf{B} \cdot \mathbf{FA} - \mathbf{A} \cdot \mathbf{FB} = \mathbf{B} \cdot \mathbf{CDA} - \mathbf{A} \cdot \mathbf{CDB}$ , a vector in the plane  $\mathbf{AB}$ . Hence the



product  $[[\mathbf{AB}][\mathbf{CD}]]$  is a vector along the line of intersection of the planes of  $\mathbf{AB}$  and  $\mathbf{CD}$ ; this is also evident geometrically.

*Ex. 10, p. 20.* When  $\mathbf{A}$  has a constant direction.

*Ex. 11, p. 20.* Since  $\oint d\mathbf{S} = 0$  for any closed surface, the integral over the curved surface is equal and opposite to the integral over the plane surface bounded by the curve. But this has the direction of the normal to the plane, which is also the direction of the binormals of the boundary curve.

*Ex. 12, p. 23.* The level surfaces of the function  $r(x, y, z)$  are spheres, so that the gradient has the radial direction. Proceeding by  $d\mathbf{r}$  in the direction normal to the sphere, the function changes by  $dr = d|\mathbf{r}|$ —hence the magnitude of the gradient is unity, and  $\text{grad } r$  represents a unit vector in the direction of  $\mathbf{r}$ . We designate it by  $\mathbf{r}_0$ .

$$\text{Ex. 13, p. 23. } \text{grad } f = \frac{df}{dr} \text{grad } r = \frac{df}{dr} \mathbf{r}_0 = \frac{1}{r} \frac{df}{dr} \mathbf{r}.$$

*Ex. 14, p. 25.* For a geometric derivation choose as volume element a portion of the spherical shell of radii  $r, r + dr$ , which subtends a solid angle  $d\Omega$  at the centre. Since  $\mathbf{r}$  is radial, it is normal to the spherical surfaces and the surface integral becomes

$$\oint \mathbf{r} d\mathbf{S} = d\Omega(r + dr)^3 - d\Omega r^3 = 3d\Omega r^2 dr.$$

Since the contained volume is  $d\tau = d\Omega r^2 dr$ , the limiting value becomes  $\text{div } \mathbf{r} = 3$ . According to the formula, this is obtained immediately from

$$\frac{\partial x}{\partial x} + \frac{\partial y}{\partial y} + \frac{\partial z}{\partial z} = 3.$$

*Ex. 15, p. 25.* The value of the exports is equivalent to the gold import; the value of goods imported is equal to the gold export. The “sources” of the outwardly directed stream of gold are the places where imported goods are consumed, and a similar statement is true for the centres of production of exported wares, which are “sinks” for the corresponding flow of gold. The integral of the gold stream over the boundaries represents the balance of trade; according to Gauss’s Theorem, this is equal to the change of the gold supply within the country. By the term “gold” we mean simply any item of credit. These considerations do not include payments without recompense (interest, &c.).

*Ex. 16, p. 29.* If radius vectors are drawn from  $O$  to points on a curve,  $d\mathbf{s} = d\mathbf{r}$ , i.e.  $\oint \mathbf{r} d\mathbf{s} = \oint d\mathbf{r}^2/2 = 0$ , so that  $\text{curl } \mathbf{r} = 0$ . This results also by direct computation from  $\partial z/\partial y = 0$ , &c.

$$\text{Ex. 17, p. 37. } t \text{ grad} \cdot t = \frac{n}{\rho}, \quad t \text{ grad} \cdot b = -\tau n.$$

*Ex. 18, p. 37.* If  $x, y, z$  are the co-ordinates of a point on an axis,

$$\left. \begin{aligned} a_{11}x + a_{12}y + a_{13}z &= \lambda x \\ a_{21}x + a_{22}y + a_{23}z &= \lambda y \\ a_{31}x + a_{32}y + a_{33}z &= \lambda z \end{aligned} \right\} \quad \cdot \cdot \cdot \cdot \cdot \cdot (1)$$

These equations, homogeneous in  $x, y, z$ , are solvable only if the determinant vanishes:

$$\left| \begin{array}{ccc} a_{11} - \lambda & a_{12} & a_{13} \\ a_{21} & a_{22} - \lambda & a_{23} \\ a_{31} & a_{32} & a_{33} - \lambda \end{array} \right| = 0. \quad \cdot \cdot \cdot \cdot \cdot (2)$$

This so-called secular equation is of third degree in  $\lambda$ . If one root is known, two of the equations (1) then give values of  $y/x$  and  $z/x$ , i.e. the directions of the principal axes. If we imagine the co-ordinate system placed along the principal axes,  $\lambda$  will give the ratio of the magnitudes of the vectors  $\Phi r$  and  $r$  for the axial directions. Since this is true for any co-ordinate system, the coefficients of the secular equation must be independent of the co-ordinates. Since, in the system of the principal axes, they contain only the  $a_{ii}$ , which we denote by  $a_I, a_{II}, a_{III}$ , we have

$$a_{11} + a_{22} + a_{33} = a_I + a_{II} + a_{III},$$

$$a_{11}a_{22} + a_{22}a_{33} + a_{33}a_{11} - (a_{12}^2 + a_{13}^2 + a_{23}^2) = a_I a_{II} + a_{II} a_{III} + a_{III} a_I,$$

$$a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - (a_{11}a_{23}^2 + a_{22}a_{13}^2 + a_{33}a_{12}^2) = a_I a_{II} a_{III}.$$

An application of this tensor invariant appears on p. 176.

\* *Ex.* 19, p. 40.

(a) For a finite volume, integration of equation (58) (p. 31) gives  $\oint u dS = \int \text{grad } u d\tau$ . But  $\text{grad } ir = i$ , hence  $\oint ir \cdot dS = \int i d\tau = iV$ .

(b) We determine first the projection upon  $i$  of the vector  $v$ , which we seek, by scalar multiplication by  $i$ :

$$iv = \oint ir \cdot i dS = i \oint ir \cdot dS = i^2 V.$$

Divide the volume into cylinders parallel to  $i$ . The lateral surfaces contribute nothing, since  $i dS = 0$ , while the bases together yield  $r \cdot i dS$ , a vector parallel to  $i$ . The resultant vector is therefore parallel to  $i$ , hence  $v = iV$ .

$$\text{Ex. 20, p. 40. } \Delta \frac{1}{r} = \text{div grad } \frac{1}{r} = \text{div} \left( -\frac{r}{r^3} \right) = +\frac{3r}{r^4} \text{grad } r - \frac{1}{r^3} \text{div } r.$$

$$\text{Now grad } r = r/r, \quad \text{div } r = 3, \quad \text{so that } \Delta \left( \frac{1}{r} \right) = 0.$$

*Ex.* 21, p. 40. Cf. p. 270. *grad*

*Ex.* 22, p. 43. Cylindrical co-ordinates:

$$(\text{grad } \psi)_z = \frac{\partial \psi}{\partial z}, \quad (\text{grad } \psi)_\rho = \frac{\partial \psi}{\partial \rho}, \quad (\text{grad } \psi)_\phi = \frac{1}{\rho} \frac{\partial \psi}{\partial \phi}$$

$$\text{div } A = \frac{\partial A_z}{\partial z} + \frac{1}{\rho} \frac{\partial (\rho A_\rho)}{\partial \rho} + \frac{1}{\rho} \frac{\partial A_\phi}{\partial \phi}$$

$$(\text{curl } A)_z = \frac{1}{\rho} \frac{\partial (\rho A_\phi)}{\partial \rho} - \frac{1}{\rho} \frac{\partial A_\rho}{\partial \phi}$$

$$(\text{curl } A)_\rho = \frac{1}{\rho} \frac{\partial A_z}{\partial \phi} - \frac{\partial A_\phi}{\partial z}$$

$$(\text{curl } A)_\phi = \frac{\partial A_\rho}{\partial z} - \frac{\partial A_z}{\partial \rho}$$

$$\Delta \psi = \frac{\partial^2 \psi}{\partial z^2} + \frac{\partial^2 \psi}{\partial \rho^2} + \frac{1}{\rho} \frac{\partial \psi}{\partial \rho} + \frac{1}{\rho^2} \frac{\partial^2 \psi}{\partial \phi^2}$$

Spatial polar co-ordinates:

$$(\text{grad } \psi)_r = \frac{\partial \psi}{\partial r}, (\text{grad } \psi)_\theta = \frac{1}{r} \frac{\partial \psi}{\partial \theta}, (\text{grad } \psi)_\phi = \frac{1}{r \sin \theta} \frac{\partial \psi}{\partial \phi}$$

$$\begin{aligned} \text{div } \mathbf{A} &= \frac{1}{r^2} \frac{\partial}{\partial r} (r^2 A_r) + \frac{1}{r \sin \theta} \frac{\partial}{\partial \theta} (\sin \theta A_\theta) \\ &\quad + \frac{1}{r \sin \theta} \frac{\partial A_\phi}{\partial \phi} \end{aligned}$$

$$(\text{curl } \mathbf{A})_r = \frac{1}{r \sin \theta} \left( \frac{\partial A_\theta}{\partial \phi} - \frac{\partial}{\partial \theta} (\sin \theta A_\phi) \right)$$

$$(\text{curl } \mathbf{A})_\phi = \frac{1}{r} \frac{\partial A_r}{\partial \theta} - \frac{1}{r} \frac{\partial (r A_\theta)}{\partial r}$$

$$(\text{curl } \mathbf{A})_\theta = \frac{1}{r} \left( \frac{\partial (r A_\phi)}{\partial r} - \frac{1}{\sin \theta} \frac{\partial A_r}{\partial \phi} \right)$$

$$\begin{aligned} \Delta \psi &= \frac{\partial^2 \psi}{\partial r^2} + \frac{2}{r} \frac{\partial \psi}{\partial r} + \frac{1}{r^2} \frac{\partial^2 \psi}{\partial \theta^2} + \frac{1}{r^2} \cot \theta \frac{\partial \psi}{\partial \theta} \\ &\quad + \frac{1}{r^2 \sin^2 \theta} \frac{\partial^2 \psi}{\partial \phi^2}. \end{aligned}$$

*Ex. 22a*, p. 46. The truth of (a) is already contained in the solution of *Ex. 18*. Here, in order to show the application for arbitrarily many dimensions, we call the co-ordinates  $x_1, x_2, x_3, \dots, x_n$  instead of  $x, y, z$ . In this notation, the system of equations (1) appearing in the solution to *Ex. 18*, corresponding to the root  $\lambda^{(h)}$  of the secular equation, becomes

$$\sum_k a_{ik} x_k^{(h)} = \lambda^{(h)} x_i^{(h)}, \quad i = 1, 2, \dots, n. \quad \dots \quad (1a)$$

Likewise, for any other root  $\lambda^{(l)}$ ,

$$\sum_k a_{ik} x_k^{(l)} = \lambda^{(l)} x_i^{(l)}, \quad i = 1, 2, \dots, n. \quad \dots \quad (2a)$$

Multiply the first equation by  $x_i^{(l)}$  and the second by  $x_i^{(h)}$ , perform a summation on each and then take the difference:

$$\sum_i \sum_k x_i^{(l)} a_{ik} x_k^{(h)} - \sum_i \sum_k x_i^{(h)} a_{ik} x_k^{(l)} = (\lambda^{(h)} - \lambda^{(l)}) \sum_i x_i^{(h)} x_i^{(l)}. \quad \dots \quad (3)$$

But  $\mathbf{A}$  is assumed to be a symmetric matrix ( $a_{ik} = a_{ki}$ ), so that the double sums on the left are equal and the components of the "characteristic vectors" \* which correspond to various roots of the secular equation have the property that their scalar product vanishes. They are mutually orthogonal in  $n$ -dimensional space.

Inasmuch as the equations (1) are homogeneous, they tell us only the ratios of the  $x_i^{(h)}$ , e.g. the directions of the vectors. Their magnitudes are still at our disposal, and we may set them equal to unity ("normalize" them). The  $x_i^{(h)}$

\* German: *Eigen*(vectors).



give the direction cosines of the co-ordinate system associated with  $A$  (in three dimensions, the axes of a second-degree surface) relative to an arbitrarily-chosen system. If these  $n^2$  quantities are used to form a matrix in such manner that the components corresponding to a given  $\lambda^{(h)}$  are placed in one column, then this matrix  $S$  will be orthogonal and the matrix  $S^{-1}$  will be its transpose. If the components of any vector  $r$  are  $x_k^*$  in the characteristic system, then the components in any arbitrary system are obtained, according to p. 34, by taking

$$r = Sr^*.$$

Multiply  $r$  by matrix  $A$ , giving a transformed vector

$$r^1 = Ar = ASr^*$$

whose components are given in the arbitrary system. According to p. 34, to obtain from this the components in the eigensystem we must take

$$r^{1*} = S^{-1}r^1 = S^{-1}ASr^*.$$

When this is calculated out,  $S^{-1}AS$  is found to be a diagonal matrix having  $\lambda^{(1)}$ ,  $\lambda^{(2)}$ , . . . as diagonal elements. This must be the case, since in the eigensystem we have simply  $x_1^{1*} = \lambda^{(1)}x_1^*$ ,  $x_2^{2*} = \lambda^{(2)}x_2^*$ , etc.

*Ex. 23, p. 57.* In the real form,  $\overline{f(t)} = 1/2$ ,

$$b_n = \frac{\omega}{\pi} \int_{\pi/\omega}^{2\pi/\omega} 1 \cdot \cos n\omega t dt = 0, \quad c_n = \frac{\omega}{\pi} \int_{\pi/\omega}^{2\pi/\omega} \sin n\omega t dt$$

or 
$$c_n = \frac{-\omega}{\pi n\omega} [\cos n\omega t]_{\pi/\omega}^{2\pi/\omega} = -\frac{2}{n\pi} \text{ for } n \text{ odd,}$$

$$= 0 \text{ for } n \text{ even.}$$

*Ex. 24, p. 57.* By equation (25) (p. 56) the coefficient  $A_s$  for the interval from  $s$  to  $s + ds$  is given by

$$A_s = \int_0^x e^{-ax} \cos \omega x e^{-2\pi i s x} dx,$$

where the variable of integration is denoted by  $x$ , and  $f(x)$  is understood to be 0 for negative values of  $x$ . This yields

$$A_s = \frac{1}{2} \int_0^\infty e^{-ax} e^{i\omega x - 2\pi i s x} dx + \frac{1}{2} \int_0^\infty e^{-ax} e^{-i\omega x - 2\pi i s x} dx$$

$$= \frac{\frac{1}{2}}{\alpha - i(\omega - 2\pi s)} + \frac{\frac{1}{2}}{\alpha + i(\omega + 2\pi s)}.$$

In practical cases  $\alpha \ll \omega$ , so that it is only when  $2\pi s$  is in the neighbourhood of  $\pm\omega$  that  $A_s$  assumes values appreciably different from zero. The conjugate complex value is obtained for  $A_{-s}$ . As may be seen by combining the two forms and passing to real values, the amplitude of the vibration of frequency  $s$  turns out to be twice the modulus of  $A_s$  (or  $A_{-s}$ ). The square of the amplitude, which

determines the intensity in the case of light vibrations, has the following value when the second integral is neglected:

$$I_s = \frac{1}{\alpha^2 + 4\pi^2(v - s)^2},$$

or for

$$v = s, \quad I_{\max.} = \frac{1}{\alpha^2}$$

and

$$I_{1/2} = \frac{1}{2} I_{\max.} \quad \text{for} \quad v - s_{1/2} = \frac{\alpha}{2\pi}.$$

*Ex. 25, p. 61.* The extreme cases are a "figure eight" closed curve for  $\delta = 0$  and a parabolic arc for  $\delta = \pi/2$ .

*Ex. 26, p. 64.* If the polar angle  $\phi$  does not enter,

$$\frac{\partial^2 u}{\partial r^2} + \frac{1}{r} \frac{\partial u}{\partial r} = \frac{1}{c^2} \frac{\partial^2 u}{\partial t^2}.$$

Setting  $u = v(r)e^{i\omega t}$ , we obtain

$$\frac{d^2 v}{dr^2} + \frac{1}{r} \frac{dv}{dr} + \frac{\omega^2}{c^2} v = 0.$$

If we put  $x = r\omega/c$ , we obtain the differential equation of cylinder functions of zero order:

$$\frac{d^2 v}{dx^2} + \frac{1}{x} \frac{dv}{dx} + v = 0.$$

This equation has two distinct integrals, the Bessel Function  $J_0(x) = J_0(r\omega/c)$ , which remains finite at the origin, and the Neumann Function which becomes logarithmically infinite there. The general solution is the sum of each multiplied by an arbitrary constant. Tables of these functions may be found in the Jahnke-Emde collection (see reference on p. 868).

*Ex. 27, p. 66.* Integration of the differential equation (54) (p. 66) under the condition  $1/V = kr$  yields the law of dispersion

$$v = \frac{v}{\sqrt{k(v^2 - C^2)}}.$$

*Ex. 28, p. 71.* Cf. p. 202, formula (32) *et seq.*

*Ex. 29, p. 74.*

$$\begin{aligned} \int_0^{2\pi} \frac{d\phi}{1 + \epsilon \cos \phi} &= \int_0^{2\pi} \frac{d\phi}{1 + \frac{\epsilon}{2}(e^{i\phi} + e^{-i\phi})} = \frac{1}{i} \oint_{\text{Unit Circle}} \frac{dz}{z + \frac{\epsilon z}{2} \left(z + \frac{1}{z}\right)} \\ &= \frac{2}{\epsilon i} \oint \frac{dz}{z^2 + \frac{2}{\epsilon} z + 1}. \end{aligned}$$

Assuming  $\epsilon < 1$ , the roots of the denominator are

$$z_1 = -\frac{1}{\epsilon} + \sqrt{\frac{1}{\epsilon^2} - 1}, \quad z_2 = -\frac{1}{\epsilon} - \sqrt{\frac{1}{\epsilon^2} - 1}.$$

The point  $z_1$  lies within the path of integration. Decomposing the denominator into partial fractions,

$$\frac{1}{(z - z_1)(z - z_2)} = \frac{1}{z_1 - z_2} \left\{ \frac{1}{z - z_1} - \frac{1}{z - z_2} \right\},$$

and the coefficient of  $1/(z - z_1)$ , the residue, becomes

$$\frac{1}{2\sqrt{\frac{1}{\epsilon^2} - 1}},$$

so that 
$$\int_0^{2\pi} \frac{d\phi}{1 + \epsilon \cos \phi} = \frac{2\pi i \cdot 2}{\epsilon i \cdot 2 \sqrt{\frac{1}{\epsilon^2} - 1}} = \frac{2\pi}{\sqrt{1 - \epsilon^2}}.$$

*Ex. 30, p. 78.* If  $F = (Fy, y')$ , the Euler-Lagrange equation gives

$$\frac{\partial^2 F}{\partial y' \partial y} y' + \frac{\partial^2 F}{\partial y'^2} y'' - \frac{\partial F}{\partial y} = 0.$$

Multiplying by  $y'$  and adding and subtracting the expression  $y''(\partial F/\partial y')$ , we obtain

$$\frac{d}{dx} \left( y' \frac{\partial F}{\partial y'} \right) - \frac{dF}{dx} = 0,$$

from which the result follows.

*Ex. 31, p. 85.* From the equation of motion

$$m \frac{d^2 r}{dt^2} = A \sin \omega t$$

it follows that

$$\frac{dr}{dt} = -\frac{A}{m\omega} \cos \omega t + C.$$

Let  $(dr/dt)_0 = v_0$  for  $t = 0$ ; then

$$C = v_0 + \frac{A}{m\omega}.$$

Integrating again,

$$r = -\frac{A}{m\omega^2} \sin \omega t + \left( v_0 + \frac{A}{m\omega} \right) t + r_0.$$

It is to be noticed that in general there is also a uniform translatory motion superposed upon the oscillation.

*Ex. 32, p. 85.* Since  $\mathbf{F} = -mg\mathbf{k}$ , where  $\mathbf{k}$  is the unit vector directed vertically upward, we have from

$$m \frac{d^2 \mathbf{r}}{dt^2} = -mg\mathbf{k}$$

the result that the motion is independent of the mass. Integrating once,

$$\frac{d\mathbf{r}}{dt} = -g\mathbf{k}t + \mathbf{v}_0.$$



Thus the hodograph is a vertical straight line. Taking the origin at the position of the point when  $t = 0$ , a second integration gives

$$\mathbf{r} = -\frac{g}{2}t^2\mathbf{k} + t\mathbf{v}_0.$$

This shows that the orbit remains in the plane of  $\mathbf{k}$  and  $\mathbf{v}_0$ . Taking this to be the  $xz$ -plane and calling the angle between the initial velocity and the  $x$ -axis  $\alpha$ , we have

$$\begin{aligned}x &= v_0 t \cos \alpha, \\z &= v_0 t \sin \alpha - \frac{1}{2}gt^2.\end{aligned}$$

Elimination of  $t$  shows that the orbit is the parabola

$$z = x \tan \alpha - \frac{gx^2}{2v_0^2 \cos^2 \alpha}.$$

Besides  $x = 0$ , there is a point  $s$  where the trajectory crosses the  $x$ -axis. This is given by

$$\tan \alpha = \frac{sg}{2v_0^2 \cos^2 \alpha} \quad \text{or} \quad s = \frac{v_0^2}{g} \sin 2\alpha.$$

This so-called range of the projectile is a maximum for  $\alpha = \pi/4$ , and has the same value for  $\alpha = \pi/4 + \beta$  as for  $\alpha = \pi/4 - \beta$ .

*Ex. 33, p. 87.* Motion is communicated to the coin by means of friction, which depends in this instance only upon the normal pressure between the surfaces. The frictional force may be taken equal to  $\mu mg$  where  $\mu$  is the coefficient of friction. Take the  $x$ -axis in the direction of motion, the origin being at the rim of the glass. Assume the rear edge of the card to be at this point also. Then, for the coin,

$$x = a + \mu g \frac{t^2}{2},$$

where  $a$  is the radius of the glass. If  $v$  is the velocity of the card, the rear edge of the card has reached the opposite rim after  $T = 2a/v$  seconds. If the coin is to fall into the glass, it must not move farther than  $x = 2a$  in this time, so that

$$x_T = a + \frac{4\mu ga^2}{2v^2} \leq 2a$$

is the condition to be satisfied.

*Ex. 34, p. 87.* No work is done, hence the feeling of fatigue must be due to other causes. The system is analogous to an electromagnet carrying a weight. No work is done—all the energy consumed appears as heat in the windings. The fact that energy is dissipated at all has nothing to do with supporting the weight, and if the windings had zero resistance no energy would have to be supplied.

*Ex. 35, p. 88.*  $U = +kr^2/2$ , since  $\text{grad } U = +kr\mathbf{r}_0 = k\mathbf{r}$ .

*Ex. 36, p. 89.* Take the fixed axis as the  $z$ -axis. Then

$$\mathbf{F} = \mu(\mathbf{r} + \lambda\mathbf{k}) = m \frac{d^2\mathbf{r}}{dt^2}.$$

Multiplying vectorially by  $\mathbf{r}$ ,

$$\lambda\mu[\mathbf{r}\mathbf{k}] = m \left[ \mathbf{r} \frac{d^2\mathbf{r}}{dt^2} \right].$$

The projection of the directed areas on the plane normal to the  $k$ -axis is obtained by taking the  $k$ -components of the auxiliary vectors. Scalar multiplication by  $k$  yields

$$0 = \left[ r \frac{d^2 r}{dt^2} \right]_z,$$

which proves the theorem.

*Ex. 37, p. 89.* If, in equation (9), p. 83, the derivatives with respect to  $t$  are replaced by those with respect to  $\phi$  according to the rule

$$\frac{d}{dt} = \frac{d}{d\phi} \dot{\phi} = \frac{2c}{r^2} \frac{d}{d\phi},$$

a simple computation leads to

$$b_r = -\frac{4c^2}{r^2} \left[ \frac{d^2(1/r)}{d\phi^2} + \frac{1}{r} \right].$$

For  $r = p/(1 - \epsilon \cos \phi)$  we obtain  $b_r \approx 1/r^3$ .

*Ex. 38, p. 94.* In the position of equilibrium  $x_0$  we have

$$\frac{\beta}{x_0^{10}} = \frac{\alpha}{x_0^3}. \quad \dots \dots \dots (1)$$

For a displacement  $s$ , we have, to a first approximation,

$$\left. \begin{aligned} F_1 &= F_1^{(0)} + \frac{2\alpha}{x_0^3} s \\ F_2 &= F_2^{(0)} - \frac{10\beta}{x_0^{11}} s \end{aligned} \right\}, \quad \dots \dots \dots (2)$$

i.e. the total force is

$$F = \left( \frac{2\alpha}{x_0^3} - \frac{10\beta}{x_0^{11}} \right) s = -\frac{8\alpha}{x_0^3} s, \quad \dots \dots \dots (3)$$

whence

$$\omega = \sqrt{\frac{8\alpha}{x_0^3 m}}, \quad \dots \dots \dots (4)$$

$m$  being the mass of the particle.

*Ex. 39, p. 100.* The maximum of the square of the amplitude involves the maximum of the function

$$\frac{1}{m^2(\omega_0^2 - \omega^2)^2 + \beta^2\omega^2},$$

which lies at  $\omega^2 = \omega_0^2 - \beta^2/2m^2$ .

On the other hand, the kinetic energy at the origin is concerned with the maximum of the function

$$\frac{\omega^2}{m^2(\omega_0^2 - \omega^2)^2 + \beta^2\omega^2}$$

on account of the derivative  $\dot{x}$ . This maximum is at  $\omega = \omega_0$ . It is to be observed that the vibrating system is non-conservative, for the friction dissipates energy; for this reason the potential energy at the points of reversal—which is proportional to the square of the amplitude—is not equal to the kinetic energy at the origin. The mean energy is

$$\bar{E} = \frac{m}{2} a^2 (\omega_0^2 + \omega^2),$$

and this is a maximum if (neglecting small quantities)

$$\omega^2 \approx \omega_0^2 - \frac{\beta^2}{4m^2} = \omega_0^2, \text{ damped.}$$

\* *Ex. 40*, p. 103. If a quadratic term is added to the elastic restoring force  $-kx$  the vibration becomes unsymmetric, since changing the sign of  $x$  does not change that of the quadratic term, so that the force has different values at two points symmetric with respect to the origin. The differential equation becomes

$$m \frac{d^2x}{dt^2} + kx + \epsilon x^2 = F_0 \cos \omega t. \quad . \quad . \quad . \quad . \quad . \quad (1)$$

As a first approximation, we assume

$$x_0 = a \cos \omega t + b. \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

The constant  $b$  allows for the dissymmetry. We insert this expression and determine  $a$  and  $b$  in such way as to make the constant term and the fundamental vibration satisfy the differential equation. Using  $\cos^2 \omega t = \frac{1}{2}(1 + \cos 2\omega t)$ , we obtain the two equations

$$b^2 + \frac{k}{\epsilon} b + \frac{1}{2} a^2 = 0, \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

and

$$a(\omega^2 - \omega_0^2) - 2 \frac{\epsilon ab}{m} + \frac{F_0}{m} = 0. \quad . \quad . \quad . \quad . \quad . \quad (4)$$

For  $\epsilon$  small, it follows from (3) that

$$b = -\frac{\epsilon a^2}{2k}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

The other solution has  $\epsilon$  in the denominator and so is of no consequence, since we consider only small correction terms. Putting the above value into equation (4), we obtain a cubic equation quite similar to that for the symmetric case:

$$\frac{\epsilon^2}{km} a^3 + a(\omega^2 - \omega_0^2) + \frac{F_0}{m} = 0. \quad . \quad . \quad . \quad . \quad . \quad (6)$$

This may be solved graphically in the same way, and under certain conditions has three real roots, so that the "skip" phenomenon occurs here also. For further details consult G. Duffing, *Erzwungene Schwingungen bei veränderlicher Eigenfrequenz und ihre technische Bedeutung*, Sammlung Vieweg Nr. 41, 42, Braunschweig (1918).

\* *Ex. 41*, p. 107. Let  $z_0$  be the height of the point of reversal above the lowest point. Then at any height  $z$  the conservation of energy gives

$$\frac{1}{2} \left( \frac{ds}{dt} \right)^2 = g(z_0 - z), \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

or

$$\frac{ds}{dt} = \sqrt{2g(z_0 - z)}. \quad . \quad . \quad . \quad . \quad . \quad . \quad (1')$$

We seek to determine  $s$  as a function of  $z$ . Putting  $s = f(z)$ , we have  $ds/dt = f'(z) dz/dt$ , and we obtain

$$dt = \frac{1}{\sqrt{2g}} \frac{f'(z)}{\sqrt{z_0 - z}} dz. \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$



Integrating from  $z = 0$  to  $z = z_0$  corresponds to a quarter period, so that

$$\frac{T}{4} = \frac{1}{\sqrt{2g}} \cdot \int_0^{z_0} \frac{f'(z) dz}{\sqrt{z_0 - z}} \quad \dots \quad (3)$$

The integral in the right member should be independent of  $z_0$ . The difficulty is that this parameter occurs not only in the integrand but also in the upper limit. To remove this difficulty we set

$$z = yz_0, \quad \dots \quad (4)$$

whence

$$\frac{T}{4} = \frac{1}{\sqrt{2g}} \int_0^1 \frac{\sqrt{z_0} f'(yz_0) dy}{\sqrt{1-y}} = \frac{1}{\sqrt{2g}} \int_0^1 \frac{\sqrt{yz_0} f'(yz) dy}{\sqrt{(1-y)y}} \quad \dots \quad (5)$$

The result is independent of  $z_0$  if

$$\sqrt{yz_0} \cdot f'(yz_0) = \sqrt{z} \cdot f'(z) = \text{const.} = a. \quad \dots \quad (6)$$

Integration gives

$$f(z) = s = 2a\sqrt{z}. \quad \dots \quad (7)$$

This curve is a cycloid, as may be shown readily: by equation (6),

$$\frac{dz}{ds} = \sin \phi = \frac{1}{a} \sqrt{z}; \quad \dots \quad (8)$$

therefore

$$z = a^2 \sin^2 \phi = \frac{a^2}{2} (1 - \cos 2\phi); \quad \dots \quad (9)$$

$$dx = \cos \phi ds = \frac{dz \cos \phi}{\sin \phi} = \frac{2a^2 \sin \phi \cos^2 \phi}{\sin \phi} d\phi, \quad \dots \quad (10)$$

or

$$x = \frac{a^2}{2} (2\phi + \sin 2\phi). \quad \dots \quad (11)$$

(9) and (11) are the familiar parametric equations of the cycloid. This method of solution is given by Marcolongo (see reference, p. 869).

*Ex. 42, p. 107.* Denoting the coefficient of friction by  $\mu$ , the condition for equilibrium on the horizontal plane is  $mv_1^2/\rho_1 = \mu mg$ , while on the cylindrical surface  $mg = \mu mv_2^2/\rho_2$ . Elimination of  $\mu$  yields

$$v_2 = \frac{g}{v_1} \sqrt{\rho_1 \rho_2}.$$

Substituting the numbers,  $v_2 = 45$  km./hr.

*Ex. 43, p. 110.* Since the motion is such that the common centre of gravity is at rest, we locate the origin there. Let  $m_2$  be on the positive side,  $m_1$  on the negative. At any instant we have

$$m_1 x_1 + m_2 x_2 = 0, \quad \dots \quad (1)$$

and

$$x_2 - x_1 = a + x. \quad \dots \quad (2)$$

The equation of motion for  $m_2$  is

$$m_2 \frac{d^2 x_2}{dt^2} + k(x_2 - x_1 - a) = 0, \quad \dots \quad (3)$$

Eliminating  $x_1$  by using equation (1), we obtain

$$\frac{d^2x_2}{dt^2} + k \left( \frac{1}{m_1} + \frac{1}{m_2} \right) x_2 = \frac{ka}{m_2}. \quad \dots \dots \dots (4)$$

If, now, we put

$$\frac{1}{\mu} = \frac{1}{m_1} + \frac{1}{m_2}, \quad \dots \dots \dots (5)$$

we obtain the ordinary equation of vibration

$$\frac{d^2x_2}{dt^2} + \frac{k}{\mu} x_2 = \frac{ka}{m_2}, \quad \dots \dots \dots (6)$$

but with a constant term which gives the rest position of  $m_2$ , since a particular integral is

$$x_2^{(0)} = \frac{a\mu}{m_2} = \frac{m_1}{m_1 + m_2} a. \quad \dots \dots \dots (7)$$

The general integral is

$$x_2 = x_2^{(0)} + A \sin(\omega_0 t - \delta), \quad \omega_0 = \sqrt{\frac{k}{\mu}}. \quad \dots \dots \dots (8)$$

Thus, in place of the mass  $m$  we have the "reduced mass"  $\mu$ .

*Ex. 44, p. 113.* Since the kinetic energy of rotation about an axis is given by

$$\frac{1}{2} \Sigma m_i d_i^2 \dot{\phi}^2,$$

where  $d_i$  is the distance of  $m_i$  from the axis, the first quantity to be computed is the moment of inertia  $I = \Sigma m_i d_i^2$  for the sphere. In spherical co-ordinates,  $d = r \sin \theta$ , so that

$$I = \rho \int_0^r \int_0^\pi \int_0^{2\pi} r^4 \sin^3 \theta dr d\theta d\phi = \frac{\rho r^5 \cdot 8\pi}{5 \cdot 3} = \frac{2Mr^2}{5}, \quad \dots \dots (1)$$

and

$$T_{\text{rot}} = \frac{1}{2} M r^2 \dot{\phi}^2; \quad \dots \dots \dots (2)$$

also,

$$T_{\text{trans}} = \frac{1}{2} M \left( \frac{ds}{dt} \right)^2. \quad \dots \dots \dots (3)$$

For rolling without slipping, we have the relation

$$ds = r d\phi; \quad \dots \dots \dots (4)$$

then, by the conservation of energy,

$$T = \left( \frac{1}{2} M + \frac{1}{2} M \right) \left( \frac{ds}{dt} \right)^2 = Mgs \sin \alpha, \quad \dots \dots \dots (5)$$

and so

$$\left( \frac{ds}{dt} \right)^2 = \frac{10}{7} gs \sin \alpha. \quad \dots \dots \dots (6)$$

If the body were to slide without rotation, the factor would be 2 in place of 10/7.

*Ex. 45, p. 113.* Designating the velocities before impact by  $u$ , after impact by  $v$ , we have for elastic collision

$$m_1 u_1 + m_2 u_2 = m_1 v_1 + m_2 v_2 \quad (\text{Conservation of Linear Momentum}) \quad (1)$$

$$m_1 u_1^2 + m_2 u_2^2 = m_1 v_1^2 + m_2 v_2^2 \quad (\text{Conservation of Energy}) \quad \dots \dots (2)$$

from which it follows that

$$v_1 = \frac{(m_1 - m_2)u_1 + 2m_2u_2}{m_1 + m_2}, \quad v_2 = \frac{(m_2 - m_1)u_2 + 2m_1u_1}{m_1 + m_2}. \quad (3)$$

For inelastic impact, the energy equation is

$$m_1u_1^2 + m_2u_2^2 = m_1v_1^2 + m_2v_2^2 + \epsilon; \quad \dots \quad (2')$$

then

$$\left. \begin{aligned} v_1 &= \frac{u_2 + \frac{m_1}{m_2}u_1 - \sqrt{(u_1 - u_2)^2 - \epsilon\left(\frac{1}{m_1} + \frac{1}{m_2}\right)}}{1 + \frac{m_1}{m_2}}, \\ v_2 &= \frac{u_1 + \frac{m_2}{m_1}u_2 + \sqrt{(u_1 - u_2)^2 - \epsilon\left(\frac{1}{m_1} + \frac{1}{m_2}\right)}}{1 + \frac{m_2}{m_1}} \end{aligned} \right\} \dots \quad (3')$$

*Ex. 46, p. 121.* We have

$$ds^2 = R^2(d\theta^2 + \sin^2\theta d\phi^2),$$

hence

$$T = \frac{mR^2}{2}(\dot{\theta}^2 + \sin^2\theta \dot{\phi}^2),$$

and the Lagrange equations for conservative forces become

$$mR^2\ddot{\theta} - mR^2 \sin\theta \cos\theta \dot{\phi}^2 + \frac{\partial U}{\partial \theta} = 0.$$

$$mR^2 \sin^2\theta \ddot{\phi} + 2mR^2 \sin\theta \cos\theta \dot{\theta}\dot{\phi} + \frac{\partial U}{\partial \phi} = 0.$$

*Ex. 47, p. 145.* The line of action for each force and for the resultant  $\mathbf{R}$  is in the vertical direction. The force polygon degenerates into a single vertical line of length  $R$ . From an arbitrary pole  $O$  draw lines to the ends of this line. These correspond to the segments  $F_{0,1}$  and  $F_{4,0}$  in fig. 5, p. 145. Through any point on  $\mathbf{R}$  draw lines parallel to  $F_{0,1}$  and  $F_{4,0}$ , cutting the verticals drawn through the ends of the beam at points  $A$  and  $B$ . Through  $O$  draw a line parallel to  $AB$  cutting  $\mathbf{R}$  in the ratio of the load forces. If there are several loads, first find their resultant and then proceed as above.

*Ex. 48, p. 151.* On account of symmetry, the principal axes of inertia are lines through the centre of gravity parallel to the edges of the cube. The moment of inertia is the same about each of these axes, so that the ellipsoid of inertia is a sphere; hence the moment of inertia about any axis through the centre of gravity has the same value. This means that the moment of inertia which we seek is equal to that about the  $z$ -axis, or

$$I = \rho \int_{-a/2}^{+a/2} \int_{-a/2}^{+a/2} \int_{-a/2}^{+a/2} (x^2 + y^2) dx dy dz = \frac{Ma^2}{6}.$$

*Ex. 49, p. 151.* If  $I_1$  is the moment of inertia about the parallel axis through



the centre of gravity, comparison of the period of the simple pendulum with that of the compound pendulum yields

$$\frac{l}{g} = \frac{I_s + Ms^2}{Mgs}$$

$$s^2 - ls + \frac{I_s}{M} = 0.$$

This is a quadratic equation for  $s$ , for a fixed value of  $l$ . The sum of the roots is

$$s_1 + s_2 = l.$$

*Ex. 50, p. 151.* The period of a compound pendulum whose centre of gravity is at a distance  $s$  from the point of suspension is

$$T = 2\pi \sqrt{\frac{I + Ms^2}{Mgs}},$$

where  $I$  is the moment of inertia about the axis through the centre of gravity and  $M$  is the mass. Putting  $I/M = a$ , we have

$$\frac{T}{2\pi} = \frac{1}{\sqrt{g}} \sqrt{s + \frac{a}{s}},$$

whence

$$\frac{\Delta T}{2\pi} = \frac{1}{2\sqrt{g} \sqrt{s + \frac{a}{s}}} \left(1 - \frac{a}{s^2}\right) \Delta s.$$

This expression vanishes when  $a/s^2 = 1$ , i.e. when  $I = Ms^2$ . Comparing the value

$$T = 2\pi \sqrt{\frac{2s}{g}}$$

for this case with the formula for the simple pendulum

$$T = 2\pi \sqrt{\frac{l}{g}},$$

we see that the equivalent length  $l$  must equal  $2s$ .

*Ex. 51, p. 151.*

$$T = \frac{1}{2}\omega^2 I + \frac{1}{2}Mv^2. \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

$$I = a_1^2 M_1 + a_2^2 M_2, \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

where  $a_1$  and  $a_2$  are the respective distances of  $M_1$  and  $M_2$  from the centre of gravity. Then

$$a_1 = \frac{M_2}{M_1 + M_2} a, \quad a_2 = \frac{M_1}{M_1 + M_2} a, \quad . \quad . \quad . \quad . \quad . \quad (3)$$

so that

$$I = \mu a^2, \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

where

$$\frac{1}{\mu} = \frac{1}{M_1} + \frac{1}{M_2}.$$

*Ex. 52, p. 160.* According to fig. 8, p. 156, the angle between  $k$  and  $k'$  is equal

to  $(\phi - \delta)$ . Then, from the magnitudes of the product vectors of equation (54), p. 160, we have  $mga \sin \delta = \bar{P} \omega \sin(\phi - \delta)$ . Solved for  $\delta$ , this is

$$\tan \delta = \frac{\sin \phi}{\frac{mga}{\bar{P} \omega} + \cos \phi}.$$

\* *Ex. 53*, p. 160. The deviation of projectiles is a very complicated problem. It is treated in detail by O. v. Eberhard in Auerbach-Hort, *Handbuch der physikalischen und technischen Mechanik*, Vol. 2. The principal effects are due to gyroscopic action. For projectiles of usual form, the effect of the air resistance is to raise the front of the shell, i.e. there is a torque about a horizontal axis normal to the axis of the projectile. Since the axis of symmetry is the principal axis of inertia in this case, we see from the relation

$$\frac{d\mathbf{P}}{dt} = \frac{d_r \mathbf{P}}{dt} + [\omega \mathbf{P}]$$

that the change in  $\mathbf{P}$ , and hence the change in the position of the axis in space, has the direction of the horizontal axis normal to the axis of the body. For right-handed spin, the deviation is toward the right when looking in the direction of advance of the projectile. At this stage the projectile is inclined to the trajectory and the forces exerted on it by its motion through the air make it veer to the right. However, after half a precessional cycle the body is inclined toward the left. It is often stated that the precession is so slow that the projectile has reached its goal before half a precession cycle is completed. This is not true. Actually the path of the tip of the projectile is not a circle; because of the changing inclination of its axis and the resulting shift in the point of application of the air resistance force, the point describes a type of cycloid which is always to the right of the tangent to the trajectory. The purely hydrodynamic Magnus Effect (p. 211) operates in the contrary direction, but is small compared with the gyroscopic effect.

*Ex. 54*, p. 167.

$$\frac{\Delta r}{r} = \frac{\sqrt{r'^2} - \sqrt{r^2}}{\sqrt{r^2}}.$$

$$\sqrt{r'^2} = r \left[ 1 + \frac{\partial u}{\partial x} \frac{x^2}{r^2} + \frac{1}{2} \left( \frac{\partial u}{\partial y} + \frac{\partial v}{\partial x} \right) \frac{xy}{r^2} + \dots \right],$$

so that

$$\begin{aligned} \frac{\Delta r}{r} = \frac{\partial u}{\partial x} \alpha_1^2 + \frac{1}{2} \left( \frac{\partial u}{\partial y} + \frac{\partial v}{\partial x} \right) \alpha_1 \alpha_2 + \frac{\partial v}{\partial y} \alpha_2^2 + \frac{1}{2} \left( \frac{\partial v}{\partial z} + \frac{\partial w}{\partial y} \right) \alpha_2 \alpha_3 \\ + \frac{\partial w}{\partial z} \alpha_3^2 + \frac{1}{2} \left( \frac{\partial w}{\partial x} + \frac{\partial u}{\partial z} \right) \alpha_3 \alpha_1. \end{aligned}$$

*Ex. 55*, p. 179. The moment of inertia of the cross-section is

$$I = \int_{r_i}^{r_e} \int_0^{2\pi} r^3 \sin^2 \phi \, dr \, d\phi = \frac{\pi}{4} (r_e^4 - r_i^4).$$

The depression of the middle is then given by equation (65) (p. 178):

$$h = \frac{Qa^3}{12\pi(r_e^4 - r_i^4)E}.$$

*Ex. 56, p. 179.* Take the origin at the uppermost point of the surface. At this point we have

$$\frac{\partial^2 z}{\partial x^2} = \frac{1}{\rho} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

$$\frac{\partial^2 z}{\partial y^2} = \frac{1}{\rho'} = -\frac{\sigma}{\rho} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

$$\frac{\partial z}{\partial x} = \frac{\partial z}{\partial y} = z = 0. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

Integrating equations (1) and (2), and using equation (3) yields the equation of the surface:

$$z = \frac{1}{2\rho} (x^2 - \sigma y^2). \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

The contour lines  $z = \text{const.}$  are hyperbolas whose asymptotes form an angle  $\alpha$  with the  $x$ -axis. This angle is given by

$$\tan^2 \alpha = \frac{1}{\sigma}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

*Ex. 57, p. 180.* Of an elementary hollow cylinder of radii  $r$  and  $r + dr$ , consider an element included between the polar angles  $\theta$  and  $\theta + d\theta$ . In the process of torsion, the upper face of this element is displaced a distance  $r\phi$  with respect to the lower face, which is at a distance  $l$  from it. Then, by p. 173, there will be a restoring force of amount  $d^2F = Srdrd\theta r\phi/l$ , whose moment is  $d^2M = r d^2F = Sr^3\phi drd\theta/l$ . Integrating over the entire ring, we have  $dM = 2\pi Sr^3 dr\phi/l$ , and integrating over all layers from  $r = 0$  to  $r = a$  yields  $M = (\pi Sa^4/2l)\phi = \tau\phi$ .  $\tau$  is called the *moment of torsion* of the wire.

*Ex. 58, p. 190.* Since  $z$  must vanish at the boundary at all times, the particular integral is of the form

$$z = c_{mn} \sin \frac{m\pi x}{a} \sin \frac{n\pi y}{b} e^{2\pi i \nu_{mn} t}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

The differential equation is satisfied if

$$\nu_{mn} = \frac{1}{2} \sqrt{\frac{m^2}{a^2} + \frac{n^2}{b^2}} \sqrt{\frac{P}{\rho}}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

where  $\rho$  is the density of the material of which the membrane is made. Equation (2) thus gives the possible natural frequencies. As in the case of the string, we write the solution, which must satisfy the initial conditions, as the sum (in this case, double sum) of particular integrals:

$$\begin{aligned} z = & \sum_0^\infty \sum_0^\infty A_{mn} \sin \frac{m\pi x}{a} \sin \frac{n\pi y}{b} \cos 2\pi \nu_{nm} t \\ & + \sum_0^\infty \sum_0^\infty B_{mn} \sin \frac{m\pi x}{a} \sin \frac{n\pi y}{b} \sin 2\pi \nu_{nm} t. \quad . \quad . \quad . \quad . \quad (3) \end{aligned}$$

The coefficients  $A_{mn}$  are obtained by the method given in the text:

$$\begin{aligned} A_{mn} &= \frac{4}{ab} \int_0^a \int_0^b f_0(x, y) \sin \frac{m\pi x}{a} \sin \frac{n\pi y}{b} dx dy \\ B_{mn} &= \frac{2}{\pi \nu_{mn} ab} \int_0^a \int_0^b f_0'(x, y) \sin \frac{m\pi x}{a} \sin \frac{n\pi y}{b} dx dy, \end{aligned}$$



where  $f_0(x, y)$  represents the initial form of the membrane, and  $f'_0(x, y)$  gives the initial distribution of velocities.

*Ex. 59, p. 194.* The force acting on an element of surface  $dS$  whose normal is directed outward is  $dF = -p dS$ . The minus sign is due to our convention regarding the sign of  $p$ . The resultant is

$$F = - \oint p dS = - \int \text{grad } p d\tau. \quad . . . . . (1)$$

If  $\rho g$  is the vector gravitational force acting on  $1 \text{ cm.}^3$ , we have by equation (1a) (p. 192),

$$F = - \int \rho g d\tau = - g \int \rho d\tau. \quad . . . . . (2)$$

*Ex. 60, p. 197.* Take the  $z$ -axis vertically downward. If the cross-section of the vessel is large and the orifice small, the flow may be considered stationary. Since the only velocity component is one in the direction of the  $z$ -axis, we have by equation (10a) (p. 196),

$$v \frac{dv}{dz} = g - \frac{1}{\rho} \frac{dp}{dz}, \quad . . . . . (1)$$

whence, by integration, 
$$\frac{v^2}{2} = gz - \frac{p}{\rho} + C. \quad . . . . . (2)$$

According to our conditions, we may take  $v = 0$  in the vessel itself and may assume that  $p$  is equal to the hydrostatic pressure  $\rho gz$ . Then  $C = 0$ . At  $z = h$ , i.e. at the orifice,  $p$  vanishes, so that

$$v_h = \sqrt{2gh}.$$

This is Torricelli's Theorem, according to which the fluid emerges with a velocity equal to that which it would attain by free fall through a distance  $h$ .

*Ex. 61, p. 197.* Taking the  $x$ -axis in the direction of flow, we have

$$v \frac{dv}{dx} = - \frac{1}{\rho} \frac{dp}{dx} \quad . . . . . (1)$$

and

$$\frac{p}{\rho} = \frac{p_1}{\rho_1}. \quad . . . . . (2)$$

Hence

$$v \frac{dv}{dx} = - \frac{p_1}{\rho_1} \frac{1}{p} \frac{dp}{dx}. \quad . . . . . (3)$$

Integrating,

$$\frac{v^2}{2} = - \frac{p_1}{\rho_1} \log p + C. \quad . . . . . (4)$$

We may set  $v = 0$  within the vessel; this determines

$$C = \frac{p_1}{\rho_1} \log p_1.$$

Then the velocity of efflux becomes

$$\frac{v_0^2}{2} = \frac{p_1}{\rho_1} \log \frac{p_1}{p_0}. \quad . . . . . (5)$$

For another gas under the same conditions of pressure,

$$\frac{v_0'^2}{2} = \frac{p_1}{\rho_1'} \log \frac{p_1}{p_0}. \quad . . . . . (5')$$



The expression on the left is the value of  $\cos \theta$  for  $r = r_0$ ; in the present case, then, ( $\cos \theta = 1$ ),

$$z_0 = \frac{2\alpha}{\rho g r_0}, \quad . . . . . (5)$$

where  $z_0$  is the mean height of the column. The same result is obtained more simply by noting that, since the liquid wets the tube completely, the surface film supports the weight of the ascended liquid. The total upward force at the film boundary is  $2\pi r_0 \alpha$ , while the gravitational force on the liquid in the tube is  $\pi r_0^2 g$ . Setting these quantities equal, we obtain (5).

*Ex. 64, p. 227.* The pressure is the same on both sides of a film, hence we may set  $p = 0$  in equation (110) (p. 224). But the resulting equation,

$$\frac{1}{R_1} + \frac{1}{R_2} = 0,$$

defines a family of minimal surfaces, i.e. surfaces having the smallest area for a given boundary curve. The rotational surface of this family, which is sought here, is the catenoid—a surface obtained by rotating the catenary  $y = \cosh x$  about the  $x$ -axis.

\* *Ex. 65, p. 235.* According to the results of p. 106, a pendulum swinging with small amplitude behaves like an oscillator with a "quasi-elastic binding constant"  $k = mg/l$ . By equation (18') (p. 234) the equations of motion are

$$m \frac{d^2 x'}{dt^2} + kx' = 2m\omega \sin \phi \frac{dy'}{dt} \quad . . . . . (1)$$

and 
$$m \frac{d^2 y'}{dt^2} + ky' = -2m\omega \sin \phi \frac{dx'}{dt}, \quad . . . . . (2)$$

which may be combined into one equation for the complex variable  $(x' + iy')$ :

$$\frac{d^2}{dt^2} (x' + iy') + \frac{k}{m} (x' + iy') = -2i\omega \sin \phi \frac{d}{dt} (x' + iy'). \quad . (3)$$

Introduce a system of co-ordinates which turns with respect to the  $x', y'$  system with the angular velocity  $\omega \sin \phi$  in a sense opposite to the rotation of the earth. Then we have

$$\begin{aligned} x' + iy' &= (\xi' + i\eta') e^{-i\omega \sin \phi t} \\ \frac{d}{dt} (x' + iy') &= e^{-i\omega \sin \phi t} \frac{d}{dt} (\xi' + i\eta') - i\omega \sin \phi (\xi' + i\eta') e^{-i\omega \sin \phi t}, \\ \frac{d^2}{dt^2} (x' + iy') &= e^{-i\omega \sin \phi t} \frac{d^2}{dt^2} (\xi' + i\eta') - 2i\omega \sin \phi e^{-i\omega \sin \phi t} \frac{d}{dt} (\xi' + i\eta') \\ &\quad - \omega^2 \sin^2 \phi (\xi' + i\eta') e^{-i\omega \sin \phi t}. \quad . . . . . (4) \end{aligned}$$

Neglecting  $\omega^2$ ,

$$\frac{d^2}{dt^2} (\xi' + i\eta') + \frac{k}{m} (\xi' + i\eta') = 0. \quad . . . . . (5)$$

Thus the ordinary pendulum equation obtains in the rotating system. This means that the plane of vibration of the pendulum, seen by an observer at rest with respect to the earth, appears to turn from east to west with the angular velocity  $\omega \sin \phi$ .



*Ex. 66, p. 235.* If the fundamental equation (7), p. 110, is resolved into components along the system of principal axes of inertia fixed in the body, it must be remembered that

$$\frac{d\mathbf{P}}{dt} = \frac{d_r\mathbf{P}}{dt} + [\boldsymbol{\omega}\mathbf{P}] \text{ and } \frac{d\boldsymbol{\omega}}{dt} = \frac{d_r\boldsymbol{\omega}}{dt}.$$

However, in this system  $P_{x'} = I_I \omega_{x'}$ ,  $P_{y'} = I_{II} \omega_{y'}$ ,  $P_{z'} = I_{III} \omega_{z'}$ , so that

$$M_{x'} = I_I \frac{d\omega_{x'}}{dt} - (I_{II} - I_{III}) \omega_{y'} \omega_{z'},$$

$$M_{y'} = I_{II} \frac{d\omega_{y'}}{dt} - (I_{III} - I_I) \omega_{z'} \omega_{x'},$$

$$M_{z'} = I_{III} \frac{d\omega_{z'}}{dt} - (I_I - I_{II}) \omega_{x'} \omega_{y'}.$$

These are known as the Euler equations. If  $\mathbf{M} = 0$  they are satisfied if

$$\omega_{x'} = \text{const.}, \omega_{y'} = \omega_{z'} = 0,$$

or 
$$\omega_{y'} = \text{const.}, \omega_{x'} = \omega_{z'} = 0,$$

or 
$$\omega_{z'} = \text{const.}, \omega_{y'} = \omega_{x'} = 0;$$

that is, in the absence of any torque a rigid body can continue to rotate about one of the principal axes. Stability considerations show that rotation about either the axis of greatest or of least inertia will be stable, while that about the middle one will be unstable.

*Ex. 67, p. 258.* The energy radiated in one year is

$$4\pi \times (150)^2 \times 10^{22} \times 2 \times 4.19 \times 10^7 \times 60 \times 24 \times 365 \text{ ergs};$$

dividing by  $c^2 = 9 \times 10^{20} \text{ cm.}^2/\text{sec.}^2$ , the decrease in mass is determined to be  $138 \times 10^{12}$  metric tons per year.

*Ex. 68, p. 269.* No, for a field of this kind is not irrotational; cf. fig. 19, p. 26.

*Ex. 69, p. 269.*

$$\sigma = -2.65 \times 10^{-4} \text{ e.s.u.} = -8.83 \times 10^{-14} \text{ coul./cm.}^2$$

$$\rho = 1.32 \times 10^{-9} \text{ e.s.u.} = 4.40 \times 10^{-19} \text{ coul./cm.}^3$$

*Ex. 70, p. 269.* Graphical differentiation of the curve yields  $-\mathbf{E}$ ; a second differentiation furnishes the value of  $-4\pi\rho$ . The differentiation is most simply performed with the aid of a mirror rule, which may be set very accurately perpendicular to the curve simply by observing when the reflection is a smooth continuation of the curve itself. The angle which the rule makes with the  $y$ -axis is equal to the slope of the tangent.

*Ex. 71, p. 276.* If the two plates are a distance  $d$  apart and if they have a difference of potential  $V$ , the uniform field between them will be of strength  $V/d$ . Apart from the slight non-uniformity at the edges, the external field is everywhere zero. This may be looked upon as the mutual cancellation of the oppositely directed lines of force from the two plates. Between the plates, on the other hand, the number of lines is doubled. According to the footnote on p. 266, the force exerted by one charged plate on the other is thus given by  $F = Ve/2d$ , and since  $C = e/V = S/4\pi d$ , we have  $F = 2\pi e^2/S$  or  $F = V^2 S/8\pi d^2$ .

*Ex. 72, p. 276.* Since the situation is independent of the polar angle and of the  $z$ -co-ordinate, we have, in cylindrical co-ordinates,

$$\Delta V = \frac{d^2 V}{dr^2} + \frac{1}{r} \frac{dV}{dr} = 0. \quad \dots \dots \dots (1)$$

The solution is  $V = A \log r + B, \quad \dots \dots \dots (2)$

whence  $V_a = A \log a + B, \quad V_b = A \log b + B, \quad \dots, \dots \dots (3)$

from which it follows for  $V_b = 0$  that

$$V = V_a \frac{\log \frac{r}{b}}{\log \frac{a}{b}}. \quad \dots \dots \dots (4)$$

The field is then given by

$$\mathbf{E} = \frac{-V_a}{\log \frac{a}{b}} \cdot \frac{r_0}{r} = \frac{V_a}{r} \cdot \frac{1}{\log \frac{b}{a}} r_0, \quad \dots \dots \dots (5)$$

and the charge  $e$  per unit length by

$$\oint \mathbf{E} d\mathbf{S} = 4\pi e = \frac{V_a \cdot 2\pi a}{a \log \frac{b}{a}}. \quad \dots \dots \dots (6)$$

Thus the capacity is

$$C = \frac{l}{2 \log \frac{b}{a}}.$$

The solution for the parallel plate condenser is

$$V = Ax + B, \quad \mathbf{E} = \frac{V}{d}, \quad \text{hence } C = \frac{S}{4\pi d}. \quad \dots \dots \dots (7)$$

The distortion of the field at the edges is neglected.

*Ex. 73, p. 276.* Call the radius of the sphere  $a$  and let the image point—whose charge is  $-e'$ —be a distance  $R'$  from the centre of the sphere. Then

$$V = \frac{e}{r} - \frac{e'}{r'}. \quad \dots \dots \dots (1)$$

The value of  $V$  must be zero on the sphere, i.e.

$$\frac{e'}{e} = \frac{r'}{r}. \quad \dots \dots \dots (2)$$

Introducing the angle  $\theta$  between a radius of the sphere and the line joining the external charge and the centre, we have

$$\frac{e'^2}{e^2} = \frac{r'^2}{r^2} = \frac{R'^2 + a^2 - 2aR' \cos \theta}{R^2 + a^2 - 2aR \cos \theta} = \frac{R' \frac{R'}{R} + \frac{a^2}{R'} - 2a \cos \theta}{R + \frac{a^2}{R} - 2a \cos \theta}. \quad \dots \dots \dots (3)$$





i.e. 
$$V_a - V_c = \int_a^c E_r dr = \frac{e}{K_1} \left( \frac{1}{a} - \frac{1}{b} \right) + \frac{e}{K_2} \left( \frac{1}{b} - \frac{1}{c} \right),$$

and since

$$C = \frac{e}{V_a - V_c},$$

we have

$$\frac{1}{C} = \frac{1}{K_1 a} - \frac{1}{K_2 c} + \frac{1}{b} \left( \frac{1}{K_2} - \frac{1}{K_1} \right).$$

*Ex. 77, p. 284.* We give the image a charge  $-e'$ . For the external space we may set

$$V_1 = \frac{e}{r} - \frac{e'}{r'}$$

without contradicting the relation  $\oint \mathbf{E} d\mathbf{S} = 4\pi e$ , since  $e'$  is not in this region, and a closed surface in this space does not enclose the charge  $-e'$ , which is really non-existent. But within the dielectric we must put

$$V_2 = \frac{e''}{Kr}.$$

The quantities  $e'$  and  $e''$  are at our disposal for satisfying the boundary conditions.

The prescribed discontinuity of the normal component yields  $e + e' - e'' = 0$ . The continuity of the tangential component is certainly satisfied if  $V_1 = V_2$  at all points on the surface of separation, for if we form the derivative  $\partial V / \partial s$  along this surface, it will have the same value on both sides if  $V_1 = V_2$ . This relation furnishes the second equation connecting  $e'$  and  $e''$ :

$$e - e' = \frac{e''}{K},$$

whence

$$e' = e \frac{K-1}{K+1} \text{ and } e'' = 2e \frac{K}{1+K}.$$

*Ex. 78, p. 286.* The force acting on the charge  $-e$  is  $-e\mathbf{E}$ , that on the charge  $+e$  is  $e(\mathbf{E} + ds \text{ grad} \cdot \mathbf{E})$ ; hence the force acting on the dipole is

$$\mathbf{F} = m \text{ grad} \cdot \mathbf{E}.$$

*Ex. 79, p. 290.* The density of surface charge is

$$\sigma = \frac{e}{S} = \frac{C(V_1 - V_2)}{S},$$

and the energy is

$$U = \frac{1}{2} \int (V_1 - V_2) \sigma dS = \frac{1}{2} C (V_1 - V_2)^2$$

or, denoting the potential difference by  $V$ ,

$$U = \frac{1}{2} C V^2.$$

If the charge is held constant,  $V' = V/K$ ,  $C' = KC$  and  $U' = U/K$ , so that

$$U' - U = \frac{1}{2} \left( \frac{1}{K} - 1 \right) C V^2 = \frac{1-K}{2K} \frac{S}{4\pi d} V^2, \text{ for a parallel-plate condenser.}$$

Now  $(1-K)/4\pi = -\kappa$ ,  $V/d = |\mathbf{E}_0|$ ,  $V/Kd = |\mathbf{E}|$ ,  $Sd = \tau$ , in conformity with (40), is proved. If the potential difference is held constant,  $U' = KU$ .

*Ex. 80, p. 290.* The value of the potential inside the sphere as given by equ. (25), p. 272, must be inserted in (31'), p. 287, and integrated from 0 to  $a$ . One then obtains

$$U = \frac{16\pi^2 \rho^2 a^5}{15} = \frac{3}{5} \frac{e^2}{a},$$

where  $e$  is the total charge.

*Ex. 81, p. 292.* Call the distance between the plates  $z$ , and measure  $z$  from the level of the lower plate, which is fixed in position. Then, if  $z'$  is the height of the counterweight,

$$U = U_{el} + U_{mech} = \frac{KSV^2}{2 \cdot 4\pi z} + mgz'. \quad \dots \dots (1)$$

For equilibrium and for a constant charge (in this case only is the total energy given by (1)), i.e. for

$$\frac{KSV}{4\pi z} = \text{const.}, \text{ or } -\frac{KS}{4\pi z^2} V \delta z + \frac{KS}{4\pi z} \frac{dV}{dz} \delta z = 0, \quad \dots \dots (2)$$

we have 
$$\delta U = 0 = +\frac{KSV^2}{8\pi z^2} \delta z + mg \delta z'. \quad \dots \dots (3)$$

Since  $\delta z' = -\delta z$ , we have, for  $z = a$ :

$$m = \frac{KSV^2}{8\pi ga^2}. \quad \dots \dots (4)$$

It appears strange that  $K$  occurs in the numerator, while the force of attraction between given charges diminished to  $1/K$  times its former value when a medium of dielectric constant replaces one whose value is unity. In this case, however, it is not the charge but the potential difference which is held constant. Cf. also Exercises 71 and 79.

*Ex. 82, p. 295.* If a quantity of electricity  $de$  is removed from the condenser in a time  $dt$  its potential difference changes by

$$dV = -\frac{de}{C}. \quad \dots \dots (1)$$

By Ohm's Law,

$$I = \frac{de}{dt} = \frac{V}{R}. \quad \dots \dots (2)$$

Combining,

$$dV = -\frac{V}{C} \frac{dt}{R}. \quad \dots \dots (3)$$

Integrated, this gives

$$\log V = -\frac{t}{CR} + \log V_0, \text{ or } V = V_0 e^{-t/CR}. \quad \dots \dots (4)$$

Hence

$$I = \frac{V_0}{R} e^{-t/CR}. \quad \dots \dots (5)$$

*Ex. 83, p. 295.* As in the preceding problem, the process of charging the condenser is found to proceed according to the equation

$$V = V_0(1 - e^{-t/CR}). \quad \dots \dots (1)$$

For  $t = T$  we must have  $V = V_s$ ; hence the time of charging—which, for instantaneous discharge, must be equal to the period of the pulsating current—is given by

$$T = RC \log \frac{V_0}{V_0 - V_s}.$$

*Ex. 84, p. 300.* A uniformly magnetized sphere may be replaced by a magnetic doublet at its centre. Its potential is

$$V = \frac{mr}{r^3}. \quad \dots \dots \dots (1)$$

From this,

$$\mathbf{H} = -\text{grad } V = -\frac{m}{r^3} + \frac{3}{r^3} mr \cdot \mathbf{r}, \quad \dots \dots \dots (2)$$

whence

$$H_r = -\frac{m \sin \phi}{r^3} + \frac{3m \sin \phi}{r^3} = \frac{2m \sin \phi}{r^3}, \quad \dots \dots \dots (3)$$

$$H_{\tan} = \frac{m \cos \phi}{r^3}, \quad \dots \dots \dots (4)$$

$$\tan \delta = \frac{H_r}{H_{\tan}} = 2 \tan \phi. \quad \dots \dots \dots (5)$$

*Ex. 85, p. 300.* It follows from

$$\oint \mathbf{H} ds = \frac{4\pi I}{c} = 0.4\pi I \quad (I \text{ in amperes}),$$

that a current whose density is  $3.9 \times 10^{-12}$  amp/cm.<sup>2</sup> flows into the surface of the earth from above. This current is many times the normal atmospheric current, and no adequate explanation of its presence has yet been given. Moreover, the reality of the vortical part of the earth's field is doubted.

*Ex. 86, p. 305.* From considerations of symmetry it follows that the magnetic lines of force are circles whose centres lie on the axis of the cylinder and whose planes are normal to this axis. It follows from

$$\oint \mathbf{H} ds = \frac{4\pi I}{c}$$

that, for a constant current density  $i$  inside,

$$2\pi r H = \frac{4\pi^2 r^2 i}{c}, \quad \text{i.e. } H_i = \frac{2\pi r i}{c};$$

outside,

$$2\pi r H = \frac{4\pi^2 R^2 i}{c}, \quad \text{i.e. } H_o = \frac{2\pi R^2 i}{rc} = \frac{2I}{rc},$$

where  $r$  is the distance of the field point from the axis of symmetry.

*Ex. 87, p. 310.* If the coil has  $N$  turns, each of cross-section  $S$ , the torque acting on it in its normal position in a field  $H$  when a current  $I$  flows is, by equation (29') (p. 309),

$$\frac{NSHI}{c} = qI, \quad \left( q = \frac{NSH}{c} \right). \quad \dots \dots \dots (1)$$





acting on the current-carrying coil amounts to  $M = qI_t = -q^2\dot{\phi}/R$ . The damping constant  $r$  in Exercise 87 thus has the value  $q^2/R$ .

Ex. 89, p. 316. Cf. p. 664 et seq.

\* Ex. 90, p. 316.

(a) We have

$$L_{12} = \frac{\mu}{c^2} \oint \oint \frac{ds_1 ds_2 \cos \phi}{r_{12}} = \frac{\mu}{c^2} \int ds_2 \int_0^{2\pi} \frac{a_1 \cos \phi d\phi}{\sqrt{a_1^2 + a_2^2 - 2a_1 a_2 \cos \phi + z^2}}.$$

In our case the angle between  $ds_1$  and  $ds_2$  is equal to the polar angle  $\phi$ .

Maxwell succeeded in referring this integral back to tabulated elliptic integrals. The formulæ for this conversion may be found in Grätz's *Handbuch der Elektrizität und des Magnetismus*, Bd. V, p. 55. We confine our considerations to two circles of almost equal size whose distance apart  $z$  is small. Introducing the largest and smallest distances

$$p^2 = z^2 + (a_1 - a_2)^2 \text{ and } q^2 = z^2 + (a_1 + a_2)^2$$

we have

$$q^2 - p^2 = 4a_1 a_2, \quad r_{12}^2 = p^2 + (q^2 - p^2) \sin^2 \frac{\phi}{2},$$

so that the integral becomes

$$L_{12} = \frac{\mu}{c^2} \cdot 2\pi a_2 \cdot 2 \int_0^\pi \frac{a_1 \cos \phi d\phi}{\sqrt{p^2 + (q^2 - p^2) \sin^2 \frac{\phi}{2}}} = \frac{4\pi a_2 \mu}{c^2} I.$$

Since, according to our assumptions,  $q \gg p$ , we can neglect  $p$  as compared with  $q$  for larger values of the angle  $\phi$ . This is not allowable when  $\phi$  is small, since  $q^2$  is multiplied by  $\sin^2 \phi/2$ . We thus divide the range of the integral into two parts—one for small values of  $\phi$  (0 to  $\alpha$ ), the other for large values ( $\alpha$  to  $\pi$ ):

$$\begin{aligned} I &\approx \int_0^\alpha \frac{a_1 d\phi}{\sqrt{p^2 + q^2 \frac{\phi^2}{4}}} + \int_\alpha^\pi \frac{a_1 (1 - 2 \sin^2 \frac{\phi}{2}) d\phi}{q \sin \frac{\phi}{2}} \\ &\approx \frac{2a_1}{q} \left( \log \frac{2a_1 \alpha}{p} - \log \frac{\alpha}{4} - 2 \right) = \frac{2a_1}{q} \left( \log \frac{8a_1}{p} - 2 \right), \end{aligned}$$

or, since  $z \ll a_1$ ,  $2a_1/q \approx 1$ ,

$$L_{12} \approx \frac{\mu}{c^2} \cdot 4\pi a_2 \left( \log \frac{8a_1}{\sqrt{z^2 + (a_1 - a_2)^2} - 2} \right).$$

This formula was given by Maxwell.

(b) On account of the expression

$$\frac{1}{8\pi} \int \mathbf{H}^2 d\tau$$

the electromagnetic field energy is a positive definite quantity. If we use the quadratic form of equation (13) (p. 316), its discriminant must always be positive, i.e.,

$$L_{11}L_{22} \geq L_{12}^2.$$

The result may be seen physically by considering the flux which a given coil causes to pass through itself and through the other coil. The most favourable case, e.g. when both coils are wound upon the same iron core, gives  $L_{11}L_{22} = L_{12}^2$ .

*Ex. 91, p. 319.* The familiar relation for resistances connected in parallel, viz.

$$\frac{1}{R} = \frac{1}{R_1} + \frac{1}{R_2},$$

holds also for complex resistances, i.e. reactances in the complex plane; hence

$$\frac{1}{R} = \frac{1}{i\omega L} + i\omega C,$$
$$R = \frac{i\omega L}{1 - \omega^2 LC}.$$

Thus for  $\omega = 1/\sqrt{LC}$  (resonance) we have  $R = \infty$ .

*Ex. 92, p. 319.* Since with increasing field the permeability of iron  $\mu = |\mathbf{B}|/|\mathbf{H}|$  decreases, the self-induction of an iron-cored coil will also decrease with increasing current. Hence the natural frequency of the circuit increases as the current grows. The details are given by Schunk and Zenneck, *Jahrbuch der drahtlosen Telegraphie*, 19, p. 170 (1922).

*Ex. 93, p. 332.* From  $\mathbf{S} = c[\mathbf{E}\mathbf{H}]/4\pi$  it follows that the mean energy flux in a vacuum ( $|\mathbf{E}| = |\mathbf{H}|$ ) is

$$\mathbf{S} = \frac{c}{8\pi} \mathbf{E}^2 = \frac{2 \times 4.19 \times 10^7}{60} \text{ ergs,}$$

whence

$$|\mathbf{E}| = \frac{4}{3} \sqrt{\frac{\pi \times 4.19}{2}} \cdot 10^{-2} \text{ e.s.u.} = 10 \text{ volts/cm.}$$

\* *Ex. 94, 342.* Both terms of equation (43) (p. 337) must be taken into account in the transitional region, so that

$$H = [\rho_0 r_0] e^{i\omega(t - r/c + \delta)}, \quad \dots \dots \dots (1)$$

where

$$\delta = \frac{1}{\omega} \tan^{-1} \left( - \frac{c}{\omega r} \right).$$

Hence the phase angle is

$$\phi(t, r) = \omega \left( t - \frac{r}{c} - \frac{1}{\omega} \tan^{-1} \frac{c}{\omega r} \right). \quad \dots \dots \dots (2)$$

The wave velocity is that of an observer  $\dot{r}$ , who moves in such way that  $\phi = \text{const.}$  Thus, by differentiation, we obtain

$$0 = 1 - \frac{\dot{r}}{c} + \frac{1}{1 + \frac{c^2}{\omega^2 r^2}} \cdot \frac{c\dot{r}}{\omega^2 r^2}, \quad \dots \dots \dots (3)$$

or

$$\dot{r} = \frac{c}{1 - \frac{1}{1 + \left(\frac{\omega r}{c}\right)^2}} > c. \quad \dots \dots \dots (4)$$



*Ex. 95, p. 354.* We have  $\theta = 45^\circ$ ,  $\theta'' = 28^\circ 7.5'$ ,  $I_s = 0.046 I_0$ ,  $I_p = 0.0042 I_0$ , whence  $P = 83$  per cent.

*Ex. 96, p. 354.* We perform the calculation for normal incidence; the procedure is entirely similar for oblique incidence. In a vacuum,  $|\mathbf{H}| = |\mathbf{E}|$ , in the dielectric  $|\mathbf{H}| = n|\mathbf{E}|$ . Also

$$|\mathbf{E}'| = \frac{n-1}{n+1} \mathbf{E}, \quad \mathbf{E}'' = \frac{2}{n+1} \mathbf{E},$$

so that we have (to a common factor)

$$|\mathbf{S}| = \mathbf{E}^2, \quad |\mathbf{S}'| = \left(\frac{n-1}{n+1}\right)^2 \mathbf{E}^2, \quad |\mathbf{S}''| = \left(\frac{2}{n+1}\right)^2 n \mathbf{E}^2,$$

from which we see that  $\mathbf{S} = \mathbf{S}' + \mathbf{S}''$ . It is to be noticed that this balance is not obtained by simply setting  $|\mathbf{S}|$  proportional to  $A^2$  in each medium, since the factor of proportionality contains  $n$ .

*Ex. 97, p. 356.*  $\tan \frac{\delta}{2} = \frac{\cos \theta \sqrt{\sin^2 \theta - n^2}}{\sin^2 \theta}$  must equal unity. In this formula

$n$  signifies the index of refraction for the ray emerging from the medium into air, i.e.  $n < 1$ ; if we put  $n = 1/n'$ ,  $n'$  is the usual index of refraction of the material. We seek the smallest value of  $n'$  for which  $\tan(\delta/2)$  can still be equal to unity. This extremal problem for  $n = f(\sin \theta)$  then yields the equation

$$n^2 = \sin^2 \theta - \frac{\sin^4 \theta}{\cos^2 \theta} = x - \frac{x^2}{1-x},$$

which is an extremum for  $n' = 2.41$ .

For two reflections

$$\tan \frac{\delta}{2} = \tan \frac{\pi}{8} = \sqrt{2} - 1 = \frac{\cos \theta \sqrt{\sin^2 \theta - n^2}}{\sin^2 \theta}.$$

For  $n = 2/3$ , this gives the two angles

$$\theta_1 = 50^\circ 20' \quad \text{and} \quad \theta_2 = 53^\circ 10'.$$

\* *Ex. 98, p. 375.* It follows from considerations of symmetry that the curves of equal difference of phase are circles. The position of the planes of vibration within the crystal may be found as follows. The wave surface corresponding to the ordinary ray is a sphere, and the same is true of the normal surface; hence in equation (26) of the text (p. 365) we put  $v_{II} = v_{III}$  for a uniaxial crystal. If we take the  $x$ -axis along the optic axis, we have  $D_x = 0$  for  $v = v_{II}$ , i.e.  $\mathbf{D}$  is normal to the axis. But since we always have  $v = v_{II}$  for the ordinary ray, the plane of  $\mathbf{D}$  is always normal to the axis. In addition,  $\mathbf{D}$  is in a plane which is normal to  $\mathbf{s}$ , i.e. it is determined by the line of intersection of a plane perpendicular to  $\mathbf{s}$  and one perpendicular to the axis. Thus, for all incident rays corresponding to a given plane of incidence,  $\mathbf{D}$  is in the same direction, viz. normal to the plane of incidence. Now, if the plane of incidence is normal to the direction of transmission of the first Nicol prism, only the ordinary ray will be present in the crystal; if the incident plane is parallel to the direction of transmission, the ordinary ray will be absent. Hence the isogyre is a right-angled cross.

*Ex. 99, p. 390.* If the index diagram be drawn as in fig. 4, p. 383, which corresponds to this case, the amplitude of the direct image will be  $A/2$ , where  $A$  is the amplitude of the incident light. The amplitude of the first order diffraction

pattern will be  $A/\pi$ ; the orders 2, 4, . . .  $2n$  yield zero, since in these instances the vector diagram is closed for any grating space. The third order, on the other hand, yields  $A/3\pi$ —in general, the odd orders have amplitudes  $A/(2n+1)\pi$ . The corresponding intensities are

$$I_0 = \frac{A^2}{4}, \quad I_1 = \frac{A^2}{\pi^2}, \quad I_3 = \frac{A^2}{9\pi^2}, \quad \dots \quad I_{2n+1} = \frac{A^2}{(2n+1)^2\pi^2}.$$

The combined intensity of the direct image and the diffraction patterns on both sides is

$$I = \frac{A^2}{4} + \frac{2A^2}{\pi^2} \left( 1 + \frac{1}{9} + \frac{1}{25} + \dots + \frac{1}{(2n+1)^2} \right) = A^2 \left( \frac{1}{4} + \frac{2\pi^2}{\pi^2 8} \right) = \frac{A^2}{2}.$$

Since the intensity of the incident light is given by  $A^2$ , and since half the energy is absorbed, the energy balance is maintained. Thus all the diffraction patterns together have the same intensity as the direct image.

*Ex. 100, p. 390.* The distance  $b$  between the slits is large compared with the slit width  $a$ ; as a result, the diffraction pattern of the single slit is crossed by maxima and minima whose separation is given by  $(\alpha - \alpha_0)b = h_1\lambda$ . For normal incidence,  $\alpha = h_1\lambda/b$ ; for almost normal incidence,  $\alpha' = h_2\lambda/b + \epsilon$ . The best visibility of the fringes is obtained for

$$\alpha = \alpha', \text{ i.e. } \frac{h_1\lambda}{b} = \frac{h_2\lambda}{b} + \epsilon \quad \text{or} \quad b = \frac{(h_1 - h_2)\lambda}{\epsilon}.$$

If  $b$  is changed, two adjoining distances for which the best visibility is obtained are separated by  $\Delta b = \lambda/\epsilon$ . This relation is of use in measuring the distance between two stars which are close together and in measuring stellar diameters. In the latter case  $\epsilon$  is approximately the angular distance between the centres of gravity of the two halves of the stellar disc.

*Ex. 101, p. 390.* From equation (30) (p. 387) it follows that

$$\sin \frac{\theta}{2} = \frac{m}{4.06} \sqrt{H_1^{*2} + H_2^{*2} + H_3^{*2}}.$$

We limit our considerations to the reflected rays which are normal to the axis of rotation; for these  $H_3^* = 0$ . Then for

$H_1^*$	$H_2^*$	$m$	$\theta/2$	$H_1^*$	$H_2^*$	$m$	$\theta/2$
1	0	1	14° 20'	1	1	1	20° 20'
1	0	2	29° 30'	1	1	2	44° 10'
1	0	3	47° 40'	2	1	1	33° 20'
1	0	4	80° 00'				

The ray reflected from the plane  $H_1^* = 2, H_2^* = 0$  coincides with the second order reflection from the plane  $H_1^* = 1, H_2^* = 0$ . Similarly, the angles resulting by interchanging  $H_1^*$  and  $H_2^*$  coincide with those already contained in the above list, and so are not set down separately.

*Ex. 102, p. 403.* It follows from  $y' = f \tan \alpha$  that the diameter of the image is 1.74 cm.

*Ex. 103, p. 403.* Let the plane in question cut the axis at P and the principal plane on the object side in the line G. The point P has a corresponding image point P' related to it by the image equation, while to G there corresponds a parallel

line  $G'$  in the principal plane on the image side.  $G'$  and  $P$  then determine the image plane required.

*Ex. 104, p. 409.* Project the incident and refracted rays on the given plane and describe a unit sphere about the intersection. The normal to the incident surface, the rays themselves and their projections on each side of the refracting plane define a right spherical triangle. Calling the angle of incidence  $\alpha$ , the angle of refraction  $\alpha''$ , the corresponding angles with respect to the given plane  $\theta$  and  $\theta''$ , and the angle between the given plane and the plane of incidence  $\beta$ , the sine law of spherical trigonometry yields  $\sin \theta = \sin \alpha \sin \beta$  and  $\sin \theta'' = \sin \alpha'' \sin \beta$ , whence  $\sin \theta / \sin \theta'' = \sin \alpha / \sin \alpha'' = n$ .

\* *Ex. 105, p. 409.* To each point of the slit there corresponds a parallel beam, and each of these beams corresponds to a point in the focal plane of the camera. Only the rays coming from the centre of the slit pass through the prism in a direction perpendicular to the edges; in fact, it is usually only the path of these rays which is reproduced in diagrams. The rays coming from the upper and lower parts of the slit pass through the prism obliquely, and a computation shows that their deflection is greater than that of the central rays. The result is that the spectral lines are curved, the convex side being toward the red. The analysis is given in H. Kayser's *Handbuch der Spektroskopie*, Bd. II, p. 260, Leipzig (1900).

*Ex. 106, p. 409.* Taking account of the signs, the focal lengths of the front and back lens surfaces are

$$f_1 = -\frac{r_1}{n-1}, \quad f_1' = \frac{nr_1}{n-1}, \quad f_2 = \frac{nr_2}{n-1}, \quad f_2' = -\frac{r_2}{n-1}.$$

If  $d$  is the lens thickness, the interval  $D = -f_1' + d + f_2$ , so that

$$f = \frac{f_1 f_2}{D} = -\frac{nr_1 r_2}{(n-1)R} \quad \text{and} \quad f' = -\frac{f_1' f_2'}{D} = \frac{nr_1 r_2}{(n-1)R},$$

where  $R = n(r_2 - r_1) + (n-1)d$ .

*Ex. 107, p. 409.* Write the sine condition in the form

$$\frac{\sin u}{\sin u'} = \frac{n'}{n} \frac{y'}{y}.$$

This relation is, obviously, also valid in the region near the axis, where the sine and the angle are interchangeable. Consider a ray from an axial point, distant  $s$  from the lens, and making an angle  $u$  with the axis. Let  $h$  be the distance from the axis at which this ray meets the first surface. Then  $\sin u = h/s$ . For a very distant point,  $s$  may be replaced by  $x$ , the distance from the principal plane on the object side. Furthermore, since  $y$  and  $y'$  are small,  $y'/y$  may be replaced by the relation  $y'/y = -f/x$  which holds for points near the axis. We then have

$$\frac{h}{\sin u} = -\frac{n'}{n} f = f'.$$

One consequence of this relation is that the largest possible value of the ratio of the diameter of a beam parallel to the axis to the focal length will be 2, if the sine condition is to hold. It is more usual to specify the *relative aperture* of an objective, which is defined as the reciprocal of the above quantity. Thus the limiting value of the relative aperture will be " $f/0.5$ ".

*Ex. 108, p. 410.* Let the prism be completely filled with light and in the position of minimum deviation. According to the principles of geometrical optics, the





This is the potential energy of the sphere (cf. *Ex.* 80). If we consider a virtual expansion  $dR$ , the energy decreases by

$$\frac{3}{8} \frac{F^2}{R^2} dR.$$

This change in energy is equal to the work done by the total pressure in the virtual expansion:

$$p dv = p \cdot 4\pi R^2 dR.$$

Equating the two quantities,

$$p = \frac{3}{8} \frac{F^2}{4\pi R^4} = \frac{3 \times (9649)^2 \times 9 \times 10^{20}}{5 \times 4\pi \times 10^4} = 4 \times 10^{23} \text{ dynes/cm.}^2 \approx 4 \times 10^{17} \text{ atm.}$$

*Ex.* 111, p. 431. The electrical deviation after traversing a condenser of length  $a$  is, by equation (8),

$$h_{el} = \frac{Ea^2}{4V}. \quad \dots \dots \dots (1)$$

It follows readily from equation (5) or equation (6) (p. 427) that the slope of the tangent to the path is

$$\tan \alpha = \frac{dy}{dx} = \frac{Ea}{2V}. \quad \dots \dots \dots (2)$$

If the screen is a distance  $l$  cm. from the end of the condenser, the total deflection is

$$H_{el} = \frac{Ea^2}{4V} + \frac{Eal}{2V} = \frac{Ea}{2V} \left( \frac{a}{2} + l \right) = \tan \alpha \left( \frac{a}{2} + l \right). \quad \dots \dots (3)$$

*Ex.* 112, p. 431. As long as  $\alpha$  is small,  $\cos \alpha$  may be set equal to 1. This amounts to taking the velocity parallel to the field to be the same for all rays. Inasmuch as the projection of the orbit on a plane normal to the field is a circle, all the electrons describe helical orbits. It is readily shown that all have the same period of revolution: From equations (11) and (12) (p. 428), it follows that  $m(d^2x/dt^2) = -(e^2/mc^2)H^2x$ . This is the well-known wave equation, for which the period is given by  $T = 2\pi mc/eH$ . The speed  $v$  does not appear, so that after a given time all particles have described one revolution and have returned to the axis. The subject of geometric electron optics, which concerns itself with image formation by electrons as brought about by electric and magnetic fields, has undergone a rapid development in recent years.

*Ex.* 113, p. 455. If there are in all  $v_1$  molecules of Substance 1 present, the number of molecules per cubic centimetre is

$$N_1 = v_1 \frac{\rho_1}{m_1},$$

where  $m_1$  is the mass,  $\rho_1$  the density of Substance 1. Then we have

$$\frac{4\pi}{3} v_1 \alpha_1 = \frac{n_1^2 - 1}{n_1^2 + 2} \frac{m_1}{\rho_1} = R_1. \quad \dots \dots \dots (1)$$

Similarly, for Substance 2,

$$\frac{4\pi}{3} v_2 \alpha_2 = \frac{n_2^2 - 1}{n_2^2 + 2} \frac{m_2}{\rho_2} = R_2. \quad \dots \dots \dots (2)$$

The mixture contains a total of  $\nu = \nu_1 + \nu_2$  molecules whose average polarizability is, say,  $\bar{\alpha}$ , whence

$$\frac{4\pi}{3} \nu \bar{\alpha} = \frac{n^2 - 1}{n^2 + 2} \frac{m}{\rho} = R. \quad (3)$$

But on account of the linear superposition of the effects of the dipoles, the average polarizability is calculated according to the simple rule

$$\bar{\alpha} = \frac{\nu_1 \alpha_1 + \nu_2 \alpha_2}{\nu_1 + \nu_2}. \quad (5)$$

Substituting this in equation (3), it follows that  $R = R_1 + R_2$ .

*Ex. 114, p. 455.* The molecular refraction of water is

$$R_m = \frac{\frac{1}{\nu^3} - 1}{\frac{1}{\nu^3} + 2} \cdot 18 = \frac{7 \times 18}{34} = 3.7 \text{ cm.}^3,$$

whence

$$r = \sqrt{\frac{3 \times 3.7}{4\pi L}} = 1.13 \times 10^{-8} \text{ cm.}$$

Further, from p. 454,

$$\frac{4\pi}{3} L\alpha = R_m,$$

and, according to p. 284,  $\alpha = a^3$ , where  $a$  is the radius of a metallic sphere. Thus  $a$  is identical with the  $r$  calculated above.

*Ex. 115, p. 455.* Since  $\nu_0 = 0$ , we obtain

$$n^2 - 1 = -\frac{Ne^2/m}{2\pi\nu^2} = -\frac{C}{\nu^2}, \quad \text{i.e. } u = \frac{c}{n} = \frac{c\nu}{\sqrt{\nu^2 - C}},$$

and

$$\frac{1}{u_g} = \frac{d\left(\frac{\nu}{u}\right)}{d\nu} = \frac{1}{c} \frac{d\sqrt{\nu^2 - C}}{d\nu} = \frac{\nu}{c\sqrt{\nu^2 - C}},$$

$$u_g = \frac{c}{\nu} \sqrt{\nu^2 - C} = \frac{c^2}{u}.$$

Thus the phase velocity in the Heaviside Layer is greater than  $c$ .

*Ex. 116, p. 457.* We have

$$\frac{4\pi N p^2}{9kT} = \frac{K - 1}{K + 2} - \frac{n^2 - 1}{n^2 + 2} = 0.75.$$

(The value of  $n$  extrapolated to  $\lambda = \infty$  is about 4/3.) It then follows that  $p = 0.8 \times 10^{-18}$  e.s.u.

This value is too small; the actual value is  $1.8 \times 10^{-18}$ . The reason for this discrepancy is association taking place in the liquid state, in which case the dipoles are no longer independent. The true value is to be obtained by measurements on gases or, in other cases, on solutions in dipole-free solvents.

\* *Ex. 117, p. 483.* For the medium at rest,  $n_0 = c/v_0$ ; further, from equation (41) (p. 244), the phase velocity in the moving medium, for an observer at rest, is

$$v = \frac{v_0 + u}{1 + \frac{uv_0}{c^2}} \approx v_0 + u - \frac{uv_0^2}{c^2},$$



where  $u$  is the velocity of the medium ( $u \ll c$ ). Then

$$v = v_0 + u \left( 1 - \frac{1}{n_0^2} \right).$$

*Ex. 118, p. 494.*  $\beta$  is given by equation (6); there results  $p = 232$  atm.

*Ex. 119, p. 500.* From the definition of  $\kappa$ , this quantity for an ideal gas is given by

$$\kappa_{id} = \frac{1}{p}.$$

For a van der Waals gas,

$$\kappa_{real} = \frac{1}{\frac{RTv}{(v-b)^2} - \frac{2a}{v^2}} \approx \frac{1}{p \left( 1 + \frac{b}{v} \right) - \frac{a}{v^2}} = \frac{1}{p} \left( 1 - \frac{b}{v} + \frac{a}{pv^2} \right).$$

*Ex. 120, p. 510.* Setting the coefficient of  $\Delta v$ , as given by (34), equal to zero and combining with (26), p. 499, we have

$$\tau = \frac{27}{4} \left( 1 - \frac{1}{3\omega} \right)^3.$$

Expressing  $\omega$  in terms of  $\tau$  and substituting in the reduced equation of state, we obtain the equation of the transition curve:

$$\Pi = -27 + 24\sqrt{3}\tau^{\frac{1}{2}} - 12\tau.$$

*Ex. 121, p. 518.* By no means! Electrical energy is equivalent to mechanical work; and according to the second law one can, by expending work, transfer heat from a cooler reservoir (say a body of water) to a warmer one (a boiler), cf. p. 517. If  $T_2 = 283^\circ$  and  $T_1 = 293^\circ$ , then  $\eta = 10/293 \doteq 1/30$  and  $W \doteq Q_1/29$  or  $Q_1 = 29W$ . Thus the thermal effectiveness may be increased, theoretically, by a factor of almost 30. "Heat pumps" of this kind are coming into use; the actual efficiency is, of course, less than the figure given.

*Ex. 122, p. 524.* Starting at an arbitrary point, say  $0^\circ$  C., the entropy before mixing is

$$S_1 = 10 \int_{273}^{373} \frac{cdT}{T} + 20 \int_{273}^{288} \frac{cdT}{T} = 10 \log \frac{373}{273} + 20 \log \frac{288}{273}$$

(specific heat of water  $c = 1$ ). After mixing, the final temperature is  $43.3^\circ$  C., and so

$$S_2 = 30 \int_{273}^{316.3} \frac{cdT}{T} = 30 \log \frac{316.3}{273}.$$

In taking the difference, the initial temperature 273 drops out—as it must do—and we have

$$S_2 - S_1 = 30 \log 316.3 - 10 \log 373 - 20 \log 288 = 0.23 \text{ cal./deg.}$$

*Ex. 123, p. 524.* In one point only. If the two curves could intersect in two points, they would enclose a surface which represents work done. One could thus construct a *perpetuum mobile* of the second kind by permitting work to be done along the isothermal—the energy being taken from a boiler—and allowing the system to return to its initial configuration by moving along the adiabatic.

*Ex.* 124, p. 524. The areas under the two arcs must be equal, for it must be impossible to construct a *perpetuum mobile* of the second kind by traversing this part of the curve.

*Ex.* 125, p. 528. We have

$$\begin{aligned} T dS &= dU + \Phi de = \left(\frac{\partial U}{\partial T}\right)_e dT + \left(\frac{\partial U}{\partial e}\right)_T de + \Phi de \\ &= T \left[ \left(\frac{\partial S}{\partial e}\right)_T de + \left(\frac{\partial S}{\partial T}\right)_e dT \right] \dots \dots \dots (1) \end{aligned}$$

Comparing coefficients,

$$\left(\frac{\partial S}{\partial e}\right)_T = \frac{1}{T} \left[ \left(\frac{\partial U}{\partial e}\right)_T + \Phi \right], \dots \dots \dots (2)$$

$$\left(\frac{\partial S}{\partial T}\right)_e = \frac{1}{T} \left(\frac{\partial U}{\partial T}\right)_e \dots \dots \dots (3)$$

Differentiating equation (2) with respect to  $T$ , equation (3) with respect to  $e$ , we have

$$\frac{1}{T} \left[ \frac{\partial^2 U}{\partial e \partial T} + \left(\frac{\partial \Phi}{\partial T}\right)_e \right] - \frac{1}{T^2} \left[ \left(\frac{\partial U}{\partial e}\right)_T + \Phi \right] = \frac{1}{T} \frac{\partial^2 U}{\partial T \partial e}, \dots \dots (4)$$

$$\text{whence} \quad \Phi = - \left(\frac{\partial U}{\partial e}\right)_T + T \left(\frac{\partial \Phi}{\partial T}\right)_e \text{ or } \Phi - T \left(\frac{\partial \Phi}{\partial T}\right)_e = - \left(\frac{\partial U}{\partial e}\right)_T \dots \dots (5)$$

If the Faraday equivalent charge  $F$  has passed, one gramme equivalent has been converted and the energy difference is equal to the heat of reaction per mol divided by the valence  $Z$ , i.e. equal to  $-q_0^-/Z$ . Then

$$\left(\frac{\partial U}{\partial e}\right)_T = -\frac{q_0^-}{ZF}$$

$$\text{and} \quad \Phi - T \left(\frac{\partial \Phi}{\partial T}\right)_e = \frac{q_0^-}{ZF}, \dots \dots \dots (6)$$

which is Helmholtz's equation for the galvanic cell. Observe that  $ZF\Phi$  corresponds to the free energy.

*Ex.* 126, p. 528. It follows from the first law, at constant pressure,

$$\delta q = du + p dv = d(u + pv) = dh = \left(\frac{\partial h}{\partial T}\right)_p dT,$$

$$\text{whence} \quad c_p = \left(\frac{\partial h}{\partial T}\right)_p.$$

By analogy with equations (34) to (37) on pp. 527-528 it follows that

$$\begin{aligned} ds &= \left(\frac{\partial s}{\partial T}\right)_p dT + \left(\frac{\partial s}{\partial p}\right)_T dp = \frac{1}{T} (du + p dv) = \frac{1}{T} (dh - v dp) \\ &= \frac{1}{T} \left\{ \left(\frac{\partial h}{\partial T}\right)_p dT + \left[ \left(\frac{\partial h}{\partial p}\right)_T - v \right] dp \right\}. \end{aligned}$$

Through comparison and "crosswise differentiation" there results

$$\left(\frac{\partial h}{\partial p}\right)_T = -T^2 \left[\frac{\partial(v/T)}{\partial T}\right]_p.$$

*Ex. 127, p. 530.* In equation (38) (p. 528), replace  $p$  by  $H$  and  $v$  by  $-l$ . The result is

$$c_H - c_l = -T \left(\frac{\partial H}{\partial T}\right)_l \left(\frac{\partial l}{\partial T}\right)_H,$$

and with the equation of state  $l = (C/T)H$ , we get

$$c_H - c_l = \frac{lH}{T} = \chi \frac{H^2}{T},$$

where  $\chi$  = susceptibility per mol and  $c_l$  is the ordinary specific heat.

\* *Ex. 128, p. 584.* For  $T = 0$ , i.e. for  $x \rightarrow \infty$ ,  $x \tanh x - \log \cosh x$  becomes equal to 2, and so  $s_m = -R \log 2$ . Statistically,  $W_{\text{mag}} = 1$ ,  $s_{\text{mag}} = 0$ .

$$W_{\text{unmag}} = \frac{L!}{\left(\frac{L}{2}\right)! \left(\frac{L}{2}\right)!} = \frac{L^L}{\left(\frac{L}{2}\right)^{L/2} \left(\frac{L}{2}\right)^{L/2}} = \frac{L^L}{\left(\frac{L}{2}\right)^L},$$

so that

$$s_{\text{unmag}} = kL \log 2 = R \log 2.$$

Thus the entropy, in agreement with the above, is an amount  $R \log 2$  greater in the unmagnetized state than in the magnetized state.

\* *Ex. 129, p. 608.* The energy levels of the rotator are given on p. 681. In the old quantum mechanics,

$$u_m = \frac{m^2 h^2}{8\pi^2 I} \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad (1)$$

Of a total number  $L$  of oscillators, the number in the  $m$ th quantum state is given by the Boltzmann Principle as

$$Z_m = \frac{Le^{-m^2 h^2 / 8\pi^2 I k T}}{\sum_m e^{-m^2 h^2 / 8\pi^2 I k T}} = \frac{Le^{-m^2 x}}{\sum_m e^{-m^2 x}} \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad (2)$$

where  $x = h^2 / 8\pi^2 I k T$ . Remembering that there are two rotational degrees of freedom, i.e. two axes normal to the axis of figure, the total energy is

$$U_r = 2LkT \frac{\sum_m m^2 x e^{-m^2 x}}{\sum_m e^{-m^2 x}} \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad (3)$$

Differentiating with respect to  $T$ , we obtain—after some computation—the rotational part of the specific heat:

$$c_r = 2Rx^2 \frac{d^2(\log \sum_m e^{-m^2 x})}{dx^2} \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad \cdot \quad (4)$$

The evaluation of the sums was first carried out by P. Ehrenfest, *Verh. d. Deutschen Phys. Ges.* 15, p. 455 (1913). It leads to qualitative agreement with the results found for  $H_2$ . Nevertheless, this simple theory is considerably in need of improvement. Every advance in the quantum theory is marked by an improvement in the theory of  $c_r$ . This was terminated by the discovery that  $H_2$  is a mix-



ture of two modifications—ortho- and para-hydrogen—one of which can possess only even, the other only odd quantum numbers. For details consult Eucken's *Lehrbuch der chemischen Physik*, Edn. 2, II, 1, p. 253 (1943).

\* *Ex.* 130, p. 624. According to p. 616, radiation incident at an angle  $\theta$  exerts a pressure amounting to  $p_\theta = (2S/c) \cos^2 \theta$  and so in an angular interval  $d\theta$ ,  $dp = 2\pi(2S/c) \cos^2 \theta \sin \theta d\theta$ . Integrating from 0 to  $\pi$  yields  $p = 8\pi S/3c$ . But  $S_\nu$  is identical with  $K_\nu$  in equation (15), and so  $p_\nu = \rho\delta/3$  and  $p = \rho/3$ . Now, at temperature  $(T + dT)$ , let a piston be pushed forward by radiation pressure of amount  $p + (dp/dT)dT$  until the volume increases by  $v$ ; then follow with adiabatic expansion that lowers the temperature from  $(T + dT)$  to  $T$ . Finally, perform a compression at temperature  $T$ . The work done is  $v(dp/dT)dT$ , while heat energy of amount

$$\left(p + \frac{dp}{dT} dT\right) v + \left(\rho + \frac{d\rho}{dT} dT\right) v$$

was supplied in the expansion. Neglecting quantities that are small compared with  $p$  and  $\rho$ , the second law yields

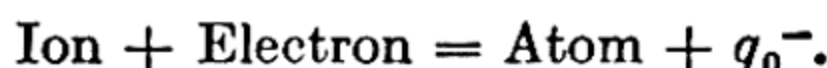
$$\frac{dp}{dT} dT = \frac{(p + \rho)dT}{T}, \quad \text{or} \quad \frac{d\rho}{dT} = \frac{4\rho}{T}.$$

Integration gives the result  $\rho = aT^4$ .

\* *Ex.* 131, p. 636. The atomic weight of electrons is 1/1840, hence the chemical constant is

$$\dot{i} = -6.48.$$

Written in exothermal form, the reaction is



The quantity  $q_0^-$  is calculated from the ionization potential  $V_i$  according to

$$q_0^- = \frac{eV_i L}{4.19 \times 10^7 \times 300} = 23,030 V_i.$$

The chemical constants of the ion and atom are equal; also, the specific heat of each of the monatomic gases is  $5R/2$ . If the degree of dissociation (ionization) is  $\alpha$ , the partial pressure of the ions is

$$p_i = \frac{\alpha}{1 + \alpha} p,$$

that of the electrons is

$$p_e = \frac{\alpha}{1 + \alpha} p,$$

and that of the atoms

$$p_a = \frac{1 + \alpha}{1 - \alpha} p,$$

since, in all,  $(1 + \alpha)N$  particles result from  $N$  neutral atoms. Then, by equation (44) (p. 541),

$$\log \frac{\alpha^2}{1 - \alpha^2} p = -\frac{q_0^-}{4.57T} + \frac{5}{2} \log T - 6.48.$$

For Ca (ionization potential 6 volts) at  $6000^\circ$  and 1 atm. we obtain  $\alpha = 8$  per cent; at  $6000^\circ$  and  $10^{-2}$  atm.,  $\alpha = 64$  per cent. Thus, observation of the relative intensity of the lines of Ca and  $\text{Ca}^+$  in stellar spectra furnishes us with the possibility of determining the conditions of temperature and pressure. This was first recognized by M. N. Saha, to whom the theory in its simple form is due. Consult *Handbuch der Astrophysik*, Vol. I, Berlin, J. Springer (1929).

\* *Ex.* 132, p. 674.

*LS* coupling:  $s_1 + s_2$  yields (a)  $S = 0$  (singlet system), (b)  $S = 1$  (triplet system);  $l_1 + l_2$  yields  $L = 1$ . (a) gives rise to the term  $^1P_1$ , and (b) to the three neighbouring terms  $^3P_0$ ,  $^3P_1$ ,  $^3P_2$ .

*jj* coupling:  $s_1 + l_1$  yields (a)  $j_1 = \frac{1}{2}$ , (b)  $j_1 = \frac{3}{2}$ ;  $s_2 + l_2$  yields  $j_2 = \frac{1}{2}$ . (a) gives rise to two neighbouring terms having  $J = 0$  and  $J = 1$ , and (b) to two terms close together having  $J = 1$  and  $J = 2$ .

Thus *LS* coupling produces a grouping of 1 and 3 terms, *jj* coupling produces two pairs of terms.

\* *Ex.* 133, p. 674. In the *longitudinal* Zeeman effect the absorption frequency for right-handed circularly polarized light is somewhat different from that for left-handed circularly polarized light. The component which rotates in the same sense as the current which would cause the magnetic field and which, by p. 663, is displaced toward higher frequencies is designated the right-handed component. In the normal effect,

$$\Delta\nu = \frac{e}{4\pi mc} H, \text{ i.e. } \nu_{0,r} = \nu_0 + \Delta\nu, \quad \nu_{0,l} = \nu_0 - \Delta\nu. \quad (1)$$

Now this displacement of the characteristic frequencies causes a change in the refractive index (p. 451); and a difference in refractive index—i.e. in phase velocity—for the right- and left-handed rays means, according to p. 65, a rotation of the plane of polarization in the direction of turning of the faster-moving wave, i.e. the wave with the smaller index of refraction. The angle of rotation is equal to half the phase difference, so that for a layer of thickness  $D$  we have

$$\alpha = \frac{\delta}{2} = \frac{2\pi\nu D}{2c} (n_l - n_r). \quad (2)$$

According to the dispersion formula in its simplest form

$$n_l - 1 = \frac{Nf_l e^2/m}{2\pi[(\nu_0 - \Delta\nu)^2 - \nu^2]}, \quad n_r - 1 = \frac{Nf_r e^2/m}{2\pi[(\nu_0 + \Delta\nu)^2 - \nu^2]}. \quad (3)$$

At higher temperatures—a condition always satisfied when investigating gases—both components are of the same intensity, so that  $f_l = f_r = f/2$ , where  $f$  is the strength of the oscillators for the uninfluenced line. At lower temperatures, at which the line spectra of crystals are observed, one of the Zeeman components may, on account of the Boltzmann distribution of orientations, be considerably weaker than the other, so that  $f_r$  and  $f_l$  will be different, and this can cause the sign of the effect to reverse, giving the so-called “paramagnetic Faraday Effect”.

Considering only the normal case where  $f_l = f_r = f/2$ , we have

$$n_l - n_r = \frac{Nfe^2}{4\pi m} \left\{ \frac{1}{[\nu_0 - \Delta\nu]^2 - \nu^2} - \frac{1}{[\nu_0 + \Delta\nu]^2 - \nu^2} \right\} = \frac{Nfe^2}{4\pi m} \cdot \frac{4\Delta\nu \cdot \nu_0}{(\nu_0^2 - \nu^2)^2} \quad (4)$$

and so

$$\alpha = \frac{Nfe^3 v v_0}{4\pi m^2 c^2 (v_0^2 - v^2)^2} HD. \quad . \quad . \quad . \quad . \quad . \quad (5)$$

Thus the rotation is proportional to  $H$ . Similarly, we find that in transverse observation, on account of the equality of the intensities of the parallel and perpendicular components, there is no linear effect but rather a small quadratic one. In the same way, the Kerr effect is found to be given as a consequence of the Stark effect.

*Ex. 134, p. 743.* According to equations (5) and (6) of the text we have, for a given velocity,  $dE/dx$  proportional to  $z^2$  but independent of  $M$ , and  $dE$  proportional to  $M$ , so that the range will be proportional to  $Mz^{-2}$ . Thus  $R_\alpha = k(4/4)$  and  $R_H = k(1.0081/1)$ , from which  $R_H = 1.008 R_\alpha$ ; i.e. for the same velocity the ranges are nearly equal. But we wish to know, from the ranges, the *energies* of the particles. Because the energy, for a given velocity, is proportional to the mass we have that

$$\frac{E_H}{E_\alpha} = \frac{1.008}{4}, \quad \text{or} \quad E_H = 3.02 \text{ Mev.}$$

This is the proton energy corresponding to a range of 14.53 cm. For heavy particles of these energies, relativistic effects may be neglected.

*Ex. 135, p. 763.* (a)  $\beta^-$  decay (neutron  $\rightarrow$  proton +  $\beta^-$ ,  $Z \rightarrow Z + 1$ ): Nucleus  $Z$  stable if  $M_{z+1} + m_{el} > M_z$  or, after addition of  $Zm_{el}$ ,  $A_{z+1} > A_z$ .

(b)  $\beta^+$  decay (proton  $\rightarrow$  neutron +  $\beta^+$ ,  $Z \rightarrow Z - 1$ ): Stability of nucleus  $Z$  if  $M_{z-1} + m_{el} > M_z$  or, after adding  $Zm_{el}$ ,  $A_{z-1} + 2m_{el} > A_z$ .

(c)  $K$ -capture (proton +  $\beta^- \rightarrow$  neutron,  $Z \rightarrow Z - 1$ ): Stability of nucleus  $Z$  if  $M_{z-1} > M_z + m_{el}$  or, after addition of  $(Z - 1)m_{el}$ ,  $A_{z-1} > A_z$ .

Stability with respect to (a) and (c)—hence also for (b)—is assured if  $A_z < A_{z-1}$  and  $A_z < A_{z+1}$ .

*Ex. 136, p. 819.* Putting  $p = \frac{1}{2}$  in (31), p. 813, and using (43), p. 818,

$$J_{3/2} = \sqrt{2/\pi x} (\sin x/x - \cos x).$$



- E. Jahnke and F. Emde, *Funktionentafeln mit Formeln und Kurven*, 3rd revised edn., Leipzig, B. G. Teubner, 1938. Reprinted, New York, Dover, 1943.
- R. Gans, *Vector Analysis with Applications to Physics*, trans. from the 6th German edn. by W. M. Deans, London, Blackie & Son, 1932.
- H. Schmidt, *Einführung in die Theorie der Wellengleichung*, Leipzig, J. A. Barth, 1932.
- L. A. Pipes, *Applied Mathematics for Engineers and Physicists*, New York and London, McGraw-Hill, 1946.
- P. M. Morse, *Vibration and Sound*, New York and London, McGraw-Hill, 1948.
- Rayleigh, *The Theory of Sound*, American edn. (rev.), New York, Dover, 1943.
- J. C. Jaeger, *An Introduction to Applied Mathematics*, Oxford, Clarendon Press, 1951.
- R. Courant and D. Hilbert, *Methods of Mathematical Physics* (translation of *Methoden der mathematischen Physik*, Berlin, Springer, 1931), New York, Interscience, 1953.
- E. M. Corson, *Introduction to Tensors, Spinors and Relativistic Wave-Equations (Relation Structure)*, London, Blackie, 1953.
- B. Higman, *Applied Group-Theoretic and Matrix Methods*, Oxford, Clarendon Press, 1955.
- T. L. Wade, *The Algebra of Vectors and Matrices*, Cambridge, Mass., Addison-Wesley, 1951.
- C. H. Page, *Physical Mathematics*, Princeton (N.J.), Van Nostrand, 1956.

## Part II

- H. Lamb, *Higher Mechanics*, 2nd edn., Cambridge, University Press, 1929.
- A. E. H. Love, *Theoretical Mechanics*, Cambridge, University Press, 1921.
- G. Hamel, *Elementare Mechanik*, 2nd edn., Leipzig, B. G. Teubner, 1922.
- E. T. Whittaker, *Analytical Dynamics*, 3rd edn., Cambridge, 1927; 4th edn., New York, Dover, 1944.
- C. Charlier, *Mechanik des Himmels*, 2 vols., 2nd edn., Berlin and Leipzig, W. de Gruyter, 1927.
- W. L. Bragg, *The Crystalline State*, Vol. I (General Survey), London, Bell, 1934.
- A. E. H. Love, *A Treatise on the Mathematical Theory of Elasticity*, 4th edn., London, Macmillan, 1927; 4th edn., New York, Dover, 1944.
- H. Lamb, *Hydrodynamics*, 6th revised edn., London, Macmillan, 1932.
- H. Lamb, *Dynamical Theory of Sound*, 2nd edn., London, Arnold, 1931.
- H. Goldstein, *Classical Mechanics*, Cambridge, Mass., Addison-Wesley, 1950.
- A. S. Eddington, *Mathematical Theory of Relativity*, Cambridge, University Press, 1923.
- A. Kopff, *The Mathematical Theory of Relativity*, trans. by H. Levy, London, Methuen, 1923.
- H. Weyl, *Space, Time and Matter*, London, Methuen, 1921.
- M. von Laue, *Die Relativitätstheorie*, 2 vols., Braunschweig, F. Vieweg, 1923.
- T. Levi-Civita, *The Absolute Differential Calculus*, London, Blackie & Son, 1927.
- G. Y. Rainich, *Mathematics of Relativity*, New York, Wiley, 1950.
- H. C. Corben and P. Stehle, *Classical Mechanics*, New York, Wiley, 1951.
- C. Moller, *The Theory of Relativity*, London, Oxford University Press, 1952.

- L. Prandtl, *Essentials of Fluid Dynamics*, London, Blackie, 1952.  
 W. H. McCrea, *Relativity Physics*, 4th edn., London, Methuen, 1954.  
 A. E. Green and W. Zerna, *Theoretical Elasticity*, Oxford, Clarendon Press, 1954.  
 L. M. Milne-Thomson, *Theoretical Hydrodynamics*, 3rd edn., London, Macmillan, 1955.  
 A. Wood, *Acoustics*, London, Blackie, 1940.  
 A. B. Wood, *A Textbook of Sound*, 3rd edn., London, Bell, 1955.  
 Shih-I Pai, *Viscous Flow Theory*, Princeton (N.J.), Van Nostrand, 1957.

### Part III

- J. H. Jeans, *The Mathematical Theory of Electricity and Magnetism*, 5th edn., Cambridge, University Press, 1925.  
 M. Abraham and R. Becker, *The Classical Theory of Electricity and Magnetism*, 2nd edn., London, Blackie & Son, 1950.  
 W. Panofsky and M. Phillips, *Classical Electricity and Magnetism*, Cambridge, Mass., Addison-Wesley, 1955.  
 R. W. Wood, *Physical Optics*, 3rd. edn., New York, Macmillan, 1934.  
 R. A. Houstoun, *Physical Optics*, London, Blackie, 1957.  
 K. Försterling, *Lehrbuch der Optik*, Leipzig, S. Hirzel, 1928.  
 J. Picht, *Optische Abbildung*, Braunschweig, F. Vieweg, 1931.  
 M. Herzberger, *Strahlenoptik*, Berlin, 1931.  
 R. W. Ditchburn, *Light*, London, Blackie, 1952.  
 Sir Edmund Whittaker, *A History of the Theories of Aether and Electricity*, London, Nelson, 1953.  
 A. R. von Hippel, *Dielectrics and Waves*, New York, Wiley, 1954.  
 E. H. Linfoot, *Recent Advances in Optics*, Oxford, Clarendon Press, 1955.  
 J. W. Smith, *Electric Dipole Moments*, London, Butterworth, 1955.  
 P. C. Shedd, *Fundamentals of Electromagnetic Waves*, New York, Prentice-Hall, 1954.  
 E. B. Wilson, etc., *Molecular Vibrations*, New York, McGraw-Hill, 1955.  
 C. Curry, *Wave Optics, Interference and Diffraction*, London, Edward Arnold, 1957.

### Part IV

- F. H. Newman, *Electrolytic Conduction*, New York, Wiley, 1931.  
 H. Falkenhagen, *Elektrolyte*, Leipzig, 1932.  
 J. J. Thomson and G. P. Thomson, *Conduction of Electricity through Gases*, 2 vols., New York, Macmillan, 1928-33.  
 F. W. Aston, *Mass Spectra and Isotopes*, London, Edward Arnold, 1933.  
 H. Huntley, *Nuclear Species*, London, Macmillan, 1954.  
 J. Meeks and J. Craggs, *The Electrical Breakdown of Gases*, Oxford, Clarendon Press, 1953.  
 K. K. Darrow, *Elementare Einführung in die Physikalische Statistik*, trans. and enlarged by E. Rabinowitsch, Leipzig, S. Hirzel, 1931; in English, *Bell System Tech. Journ.*, **8**, p. 672 (1929), reprinted in *Physical Review Supplement*, **1**, p. 90 (1929).

L. Brillouin, *Quantenstatistik*, trans. by E. Rabinowitsch, Berlin, J. Springer, 1931.

J. H. van Vleck, *The Theory of Electric and Magnetic Susceptibilities*, Oxford, Clarendon Press, 1932.

D. Schoenberg, *Superconductivity*, 2nd edn., Cambridge, University Press, 1952.

P. Debye, *The Structure of Molecules*, London, Blackie, 1932.

R. Becker and W. Döring, *Ferromagnetismus*, Berlin, Springer, 1939.

T. S. Moss, *Photoconductivity in the Elements*, London, Butterworth, 1952.

H. S. W. Massey and E. H. S. Burhop, *Electronic and Ionic Impact Phenomena*, Oxford, Clarendon Press, 1952.

### Part V

M. Planck, *Treatise on Thermodynamics*, trans. by A. Ogg, 3rd rev. edn., New York, Dover, 1943.

J. K. Roberts, *Heat and Thermodynamics*, 4th edn., London, Blackie & Son, 1951.

A. Eucken, *Fundamentals of Physical Chemistry*, trans. from the 2nd German edn. by E. R. Jette and V. La Mer, New York, McGraw-Hill, 1925; 5th German edn., Leipzig, 1942.

A. Eucken, *Lehrbuch der chemischen Physik*, 2nd edn., Leipzig, Akad. Verlagsges., 1943.

J. R. Partington, *Chemical Thermodynamics*, New York, Van Nostrand, 1924.

A. Findlay, *The Phase Rule and its Applications*, 8th rev. edn., New York, Dover, 1943.

J. H. Jeans, *The Dynamical Theory of Gases*, 4th edn., Cambridge, University Press, 1925.

L. B. Loeb, *The Kinetic Theory of Gases*, New York, McGraw-Hill, 2nd edn., 1934.

E. H. Kennard, *Kinetic Theory of Gases*, New York, McGraw-Hill, 1938.

R. A. Smith, *The Physical Principles of Thermodynamics*, London, Chapman & Hall, 1952.

A. R. Ubbelohde, *An Introduction to Thermodynamical Principles*, 2nd edn., Oxford, Clarendon Press, 1952.

N. A. Hall, *Thermodynamics of Fluid Flow*, New York, Prentice-Hall, 1951.

E. A. Guggenheim, *Thermodynamics*, 3rd edn., Amsterdam, North-Holland Publishing Co., 1957.

### Part VI

R. H. Fowler, *Statistical Mechanics*, Cambridge, 1929.

M. Planck, *The Theory of Heat* (vol. 5 of *Introduction to Theoretical Physics*), London, Macmillan, 1932.

E. O. Hirsch, *Elements of Thermodynamics and Statistical Mechanics*, Cambridge, University Press, 1951.

W. P. Allis and M. A. Herlin, *Thermodynamics and Statistical Mechanics*, New York, McGraw-Hill, 1952.



- S. Chapman and T. G. Cowling, *The Mathematical Theory of Non-Uniform Gases*, 2nd edn., Cambridge, University Press, 1952.
- D. ter Haar, *Elements of Statistical Mechanics*, New York, Rinehart, 1954.
- M. Dole, *Introduction to Statistical Thermodynamics*, New York, Prentice-Hall, 1955.
- A. H. Wilson, *Thermodynamics and Statistical Mechanics*, Cambridge, University Press, 1957.

### Part VII

- A. Sommerfeld, *Atomic Structure and Spectral Lines*, trans. from the 3rd German edn. by H. L. Brose, New York, E. P. Dutton, 1923.
- E. Schrödinger, *Collected Papers on Wave Mechanics*, London, Blackie & Son, 1928.
- L. de Broglie and L. Brillouin, *Selected Papers on Wave Mechanics*, London, Blackie & Son, 1928.
- A. Sommerfeld, *Wave Mechanics*, trans. by H. L. Brose, London, Methuen, 1930.
- P. A. M. Dirac, *Principles of Quantum Mechanics*, Oxford, Clarendon Press, 1930.
- E. Rutherford, J. Chadwick and C. Ellis, *Radiations from Radioactive Substances*, New York, Macmillan, 1930.
- G. Gamow, *The Constitution of Atomic Nuclei and Radioactivity*, Oxford, Clarendon Press, 1931.
- H. Kopfermann, *Kernmomente*, Leipzig, Akad. Verlagsges., 1940; Reprint edn., Ann Arbor, Edwards Bros., 1945.
- G. Herzberg, *Atomic Spectra and Atomic Structure*, 2nd rev. edn., New York, Prentice-Hall, 1937.
- G. Herzberg and J. Spinks, *Molecular Spectra and Molecular Structure*, New York, van Nostrand, 1950.
- F. Seitz, *Modern Theory of Solids*, New York and London, McGraw-Hill, 1940.
- F. Richtmyer, E. H. Kennard, and T. Lauritsen, *Introduction to Modern Physics*, 5th edn., New York, McGraw-Hill, 1955.
- R. A. Millikan, *Electrons (+ and -), Protons, Photons, Neutrons, Mesotrons and Cosmic Rays*, rev. edn., Chicago, University of Chicago Press, 1947.
- L. Janossy, *Cosmic Rays and Nuclear Physics*, London, Pilot Press, 1948.
- S. Fluegge and J. Mattauch, *Nuclear Physics Tables and Introduction to Nuclear Physics*, trans. by E. Gross and S. Bargmann, New York, Interscience, 1946.
- A. March, *Quantum Mechanics of Particles and Wave Fields*, New York, Wiley, 1951.
- W. V. Houston, *Principles of Quantum Mechanics*, New York, McGraw-Hill, 1952.
- A. Lande, *Quantum Mechanics*, London, Pitman, 1951.
- N. F. Ramsey, *Nuclear Moments*, New York, Wiley, 1954.
- D. J. Hughes, *Neutron Optics*, New York, Interscience, 1954.
- R. D. Evans, *The Atomic Nucleus*, New York, McGraw-Hill, 1955.
- A. E. S. Green, *Nuclear Physics*, New York, McGraw-Hill, 1955.

- R. E. Lapp and H. L. Andrews, *Nuclear Radiation Physics*, 2nd edn., London, Pitman, 1955.
- W. Heitler, *The Quantum Theory of Radiation*, Oxford, Clarendon Press, 1954.
- F. Mandl, *Quantum Mechanics*, London, Butterworth, 1954.
- O. Oldenburg, *Introduction to Atomic Physics*, 2nd edn., New York, McGraw-Hill, 1954.
- R. S. Shankland, *Atomic and Nuclear Physics*, New York, Macmillan, 1956.
- M. Born, *Atomic Physics*, 6th edn., London, Blackie, 1957.
- R. H. Aitken, *Mathematics and Wave Mechanics*, London, Heinemann, 1957.
- H. T. Flint, *Wave Mechanics*, 8th edn., London, Methuen, 1957.

### *Part VIII*

- V. K. Zworykin, G. A. Morton, E. G. Ramberg, and others, *Electron Optics and the Electron Microscope*, New York, Wiley, 1945.
- W. G. Cady, *Piezoelectricity*, New York, McGraw-Hill, 1946.
- R. Houwink, *Elastomers and Plastomers: Their Chemistry, Physics and Technology*, 3 vols, New York, Elsevier, 1950.

### *Mathematical Addendum*

- T. M. MacRobert, *Spherical Harmonics*, 2nd rev. edn., New York, Dover, 1944.
- F. E. Relton, *Applied Bessel Functions*, London, Blackie, 1946.





# INDEX

- Abaric point, 846, Ex. 74.
- Abbe's sine law, 403.
- Aberration, 245.
- Absolute electrometer, 292.
- space, 228.
- velocity, 230.
- zero, unattainable, 559.
- Absorbing media, 357.
- Absorption frequency, 453.
- Absorptive power, 620.
- Accelerated reference frame, 231.
- Acceleration, 82.
- Action, 123.
- at a distance, 286.
- variable, 129.
- Adiabatic change, 506.
- law, 212.
- lines, 507.
- principle, 662.
- Aeromechanics, 191.
- Affinor, 33 n.
- Alpha-particle, 646.
- scattering of, 641.
- Alpha-ray, 745.
- Alternating current, 316.
- — transient phenomena, 319.
- Alvarez, 772.
- Ampère's theorem, 305.
- Amplitude, 50.
- Anfahrwirbel, 211 n.
- Angular frequency, 50.
- momentum, 89.
- — total, 110.
- Anisotropic substances, 278.
- Ankylosis, 598.
- Antecedent, 35.
- "Anti-Stokes" line, 721.
- Apparent force, 231.
- Arc spectra, 674 n.
- Argand diagram, 69.
- Associated harmonics, 700, 821.
- Astigmatic difference, 409.
- Aston's isotope rule, 762.
- Atled operator, 30 n.
- Atom, model of, 647.
- structure of, 641.
- Atomic mass unit, 750.
- number, 431, 646.
- observations, inexactness of, 690.
- Avogadro's law, 495.
- number, 420, 572, 598.
- Axes, optical, 369.
- Axis of figure, 153.
- Babinet's theorem, 380.
- Balmer series of spectral lines, 649.
- Band group, 685.
- spectra, 680.
- system, 685.
- Band, 731.
- Bars, longitudinal waves in, 180.
- Beats, 57, 323.
- Bernoulli's equation, 197, 203.
- Berthelot's principle, 513.
- Bessel functions, 810.
- — asymptotic forms of, 817.
- — of second kind, 816.
- — of third kind, 817.
- — orthogonality of, 815.
- Beta-ray, 745.
- Binormal, 19.
- Biogenetics, fundamental law of, 675.
- Biot-Savart law, 110, 206, 207, 300, 303, 337, 338.
- Black body, 613, 620, 624.
- Bloch and Alvarez, 772.
- Bohr magneton, 665, 744.
- model of atom, 647.
- theory, 707.
- Boiling-point, elevation of, 546.
- Boltzmann constant, 452, 593.
- Bond, non-polar, 712.
- Born, M., 706.
- Bose-Einstein statistics, 624.
- Boundary layer, 210.
- — detachment of, 210, 211 n.
- — waves in, 345.
- Boyle and Mariotte, law of, 494, 565.

- Boyle's law, 212.  
 Brackett series of spectral lines, 650.  
 Bragg law, 388.  
 — method, 389.  
 Breakdown, electrical, 436.  
 Bremsstrahlung, 743, 779.  
 Brewster's law, 353.  
 Brownian movement, rotational, 602 n.  
 Bulk modulus, 173.  
 Bunsen's effusometer, 842, Ex. 61.
- Calculus of variations, 75.  
 Calorie, gramme, 487.  
 — large, 487.  
 — small, 487.  
 Canal rays, 440.  
 — — Doppler effect in, 441.  
 Canonical transformation, generator of, 127.  
 — transformations, 124.  
 Cantilever beam, bending of, 176.  
 Capacity, 274.  
 Carnot cycle, 514.  
 Cathode drop, 437.  
 — rays, 426, 427.  
 — — origin of, 440.  
 Cations, 419, 441.  
 Cauchy integral theorem, 72.  
 Cauchy-Riemann conditions, 71.  
 — — equations, 201.  
 Caustic, 406.  
 Cavity, 613.  
 Central axis, 144.  
 Centre of gravity, 109.  
 — of inertia, 109.  
 — of mass, motion of, 108.  
 Characteristic curve of thermionic tube, 799.  
 — functions, 186.  
 — radiation, infra-red, 607.  
 — values, 185.  
 — — of wave equation, 698.  
 Charge density, 703.  
 — surface density of, 266.  
 Chemical affinity, 553.  
 — bond, nature of, 712.  
 — constant, 541 636.  
 — — of gas, 538, 557.  
 — equilibrium, 543.  
 — valence, 716.  
 Circuits, inductively coupled, 322.  
 Circularly polarized waves, 64.  
 Circulation, 202, 203, 210, 211.  
 Classical quantum statistics, 603.
- Classical statistics, 579, 603.  
 Clausius-Clapeyron equation, 533.  
 Cloud chamber, 740.  
 — — Wilson's, 642.  
 Coefficient of emission, 619.  
 Collinear system, 403.  
 Combination lines, 660.  
 Compensator, 357.  
 Complex potential, 201, 209, 210.  
 — velocity, 203, 207, 208, 210.  
 "Complexion", 582.  
 Compressibility, isothermal, 493.  
 Compton effect, 687, 690, 691, 743.  
 Conditionally periodic systems, 132.  
 Conducting media, propagation in, 333.  
 Conduction, electrolytic, 417.  
 — metallic, 444.  
 — of electricity in gases, 425.  
 — of heat, 570, 571.  
 — — in metals, 447.  
 Conductivity, 294.  
 — electrical, 631.  
 — equivalent, 421.  
 — surface, 489.  
 — thermal, 488.  
 — thermometric, 489.  
 Conformal mapping, 69, 70.  
 Conjugate complex velocity, 201.  
 — tensors, 35.  
 Consequent, 35.  
 Conservation of energy, principle of, 88.  
 Conservative forces, 87.  
 Constrained particle, 103.  
 Contravariant components, 12.  
 Convection coefficient, Fresnel's, 477.  
 Convergence ratio, 401.  
 Co-ordinates, curvilinear orthogonal, 40.  
 Coriolis acceleration, 233.  
 Correspondence principle, 655.  
 Corresponding states, theorem of, 500.  
 Cosmic-ray showers, 781.  
 Cosmic rays, 780.  
 Coulomb, 267.  
 Coulomb's law, 266.  
 Couple, 144.  
 Coupled circuits, 322.  
 Coupling, Russell-Saunders, 669.  
 Covariant components, 12.  
 Crystal, biaxial, 373.  
 — cohesive forces in, 574.  
 — compressibility of, 577.  
 — lattice, 360.  
 — uniaxial, 370.  
 Crystals, properties of, 573.

- Cubical dilation, 166.  
 Curie and Joliot, 755, 756.  
   — constant, 458.  
   — law, 529.  
   — point, 460, 461.  
 Curl of a vector, 25.  
 Current density, 293, 726.  
   — strength, 293.  
 Curves of equal thickness, 411.  
 Cut, 73 n.  
 Cyclic variables, 126.  
 Cylinder functions. See *Bessel functions*.
- d'Alembert's principle, 113, 116.  
 Dalton's law, 496.  
 Damping, 320, 348.  
 Debye-Hückel theory of electrolytic conductivity, 421, 423.  
 Decomposition of lithium, 750.  
 Deformable solids, mechanics of, 161.  
 Degenerate system, 136.  
 Degree of degeneracy, 136.  
 Del operator, 30 n.  
 Delaunay elements, 136.  
 Density of charge, 265.  
   — of ideal gas, 580.  
 Deuteron, 653, 743, 751.  
 Dia-electric susceptibility, 280.  
 Diamagnetic susceptibility, 463.  
 Diamagnetism, 451.  
 Diatomic molecule, 585, 605.  
   — molecules, rotation spectra of, 702.  
 Dielectric constant, 267 n., 277.  
   — — electron theory of, 450.  
   — constants, principal, of crystal, 362.  
   — displacement, 277.  
   — media, 277.  
   — polarization, 451, 473.  
 Dielectrics, polarization of, 279.  
 Diffraction at circular aperture, 390.  
   — at slit, 382, 390.  
   — optical, 696.  
   — reciprocal theorems of, 379.  
   — theory of, 376.  
 Diffuse series, 660.  
 Diffusivity, 489.  
 Dilatation, Einstein time, 243.  
 Diopter, 403.  
 Dipole, 272.  
 Discord, 99.  
 Dispersion, 65, 451, 716.  
   — anomalous, 453.  
   — region of, 413.  
 Displacement current, 327, 328, 475.
- Divergence of a vector, 23.  
   — theorem, 24.  
 Domains, magnetic 462.  
 Doppler effect, 235, 245, 691.  
 Double layer potential of, 273.  
 Doublet, 273.  
   — magnetic, 298.  
 Droplets, 263.  
 Dulong-Petit law of atomic heats, 598.  
 Dyad, 35.  
 Dyadic products, 36.  
 Dynamics of rigid bodies, 143.
- Earth's horizontal magnetic field, intensity of, 297.  
   — pole, wandering of, 155.  
 Eccentric anomaly, 137.  
 Efficiency of heat engine, 517.  
 Effusometer, Bunsen's, 842.  
 Eigenfunktion, 186 n.  
 Eigenwert, 186 n.  
 Einstein-de Hass effect, 111.  
   — — — anomaly in, 669.  
   — time dilatation, 243.  
 Einstein's velocity addition theorem, 244, 245.  
 Elastic potential, 175, 176.  
 Elasticity, 161.  
 Elastodynamics, 161.  
 Elastomers, 803.  
 Elastostatics, 161.  
 Electret, 473.  
 Electrical charge as source of flux, 264.  
   — oscillator, 338.  
 Electrocaloric phenomena, 528.  
 Electrodynamics of moving bodies, 471.  
 Electrolytic theory of Arrhenius, 422.  
 Electromagnetic equations, invariant form of, 480.  
   — waves, 360.  
   — — in conducting media, 333.  
   — — in two media, 343.  
   — — propagation of, 328.  
 Electrometer, absolute, 292.  
 Electron, 426, 429.  
   — mass, variability of, 476.  
   — optics, 443, 787.  
   — revolving, 463.  
   — spinning, 669.  
 Electrons, secondary, 433.  
   — thermal emission of, 433.  
 Electrophoretic force, 423.  
 Electrostatic potential, 267.  
 Elementary cell, 574.



- Ellipsoid, Fresnel, 363.  
 — index, 363, 366.  
 — of stress, 169.  
 Ellipsoid, tensor, 363.  
 Elliptic motion, left-handed, 59.  
 — — right-handed, 59.  
 Elliptically polarized waves, 64.  
 Emissivity, 489, 619.  
 Enclosure, 613.  
 Endothermal reaction, 540.  
 Energy, 85.  
 — bands, 725.  
 — density, 175, 287, 616.  
 — distribution, 588.  
 — function of a gas, 503.  
 — inertia of, 252.  
 — internal, 527.  
 Enthalpy, 509.  
 — free, 525.  
 Entrance pupil, 785.  
 Entropy, 522, 524, 540, 556, 579.  
 — constant, 636.  
 Equation of state, 493, 527.  
 — — of ideal gas, 494.  
 — — of real gas, 497.  
 Equations of motion, canonical form, 122.  
 Equilibrium, chemical, 539.  
 — of disintegration products, 746.  
 — of sedimentation, 589.  
 — of thermodynamic systems, 531.  
 — thermodynamic, 619.  
 Equipotentials, 201, 202.  
 Ether, 259.  
 Euler formula, 50.  
 Euler-Lagrange differential equation, 75, 77.  
 Euler-Mascheroni constant, 809.  
 Eutectic points, 534.  
 Exchange energies, 712.  
 — forces, 712.  
 Excitation potential, 649.  
 Exit pupil, 785.  
 Expansion, isobaric volume coefficient of, 493.  
 External forces, 108.  
 “Falling characteristic” of arc, 440.  
 Faraday effect, 674, Ex. 133.  
 — — paramagnetic, 863, Ex. 133.  
 — equivalent charge, 554.  
 Fermat's principle, 395, 407.  
 Fermi-Dirac statistics, 625, 631.  
 Fermi gas, 769.  
 Ferromagnetic materials, 305, 306.  
 Ferromagnetic media, 343.  
 — metals, 299, 460.  
 Field equations, Hetz's solution of, 335.  
 — strength, electrical, 263.  
 Fields, quasi-stationary, 311.  
 — rapidly alternating, 327.  
 — slowly varying, 311.  
 “Fine structure”, 669.  
 First law of thermodynamics, 495.  
 Fission of heavier nuclei, 767.  
 Fizeau effect, 245.  
 Fizeau's experiment, 477.  
 Fleming's left-hand rule, 309.  
 Flow, irrotational, 196, 197.  
 — laminar, 214.  
 — rotational, 196.  
 Flux, 264.  
 Focal length, 399.  
 — — of electron lens, 763.  
 — plane, 396.  
 Focus, principal, 420.  
 Force, 84.  
 — moment of, 110.  
 Forced vibrations, 96.  
 Fortrat, 684.  
 Foucault pendulum, 235, Ex. 65.  
 Four-dimensional world, 247.  
 Four-force, 256.  
 Four-velocity, 256.  
 Fourier integral, 54, 56.  
 — series, 54, 101.  
 — — multiple, 132.  
 Fraunhofer diffraction phenomena, 382.  
 Free charge, 277.  
 — energy, 525.  
 — oscillations, 320. ✓  
 — path, 438.  
 Freezing-point, lowering of, 546.  
 Frequency, 50.  
 Fresnel diffraction phenomena, 382.  
 Fresnel's formulæ, 350, 352.  
 Frictional resistance, 94.  
 Funicular polygon, 145.  
 Galilean transformation, 229, 230.  
 Galvanometer, moving-coil, 309.  
 Gamma functions, 808.  
 — -ray, 745.  
 Gas constant, 593.  
 Gauss plane, 69.  
 Gauss's theorem, 23, 270.  
 Gay-Lussac's experiment, 504.  
 — law, 494.  
 Geiger-Müller counter, 741.

- Geiger-Nuttall formula, 747.  
 Generalized co-ordinates, 118.  
 — momentum co-ordinates, 121.  
 Geodesic, 75.  
 Geometric velocity, 245.  
 Giorgi system of units, 266.  
 Glancing angle, 389.  
 Goudsmit and Uhlenbeck, 669.  
 Gradient, 21.  
 Gravitational mass, 259.  
 Green's theorem, 270.  
 Grid, 434.  
 Ground wave, 345.  
 Group velocity, 65, 245, 694.  
 Gyrocompass, 159.  
 Gyroscope, 152.  
  
 Hahn and Strassmann, 754.  
 Half-value period, 746.  
 Hall effect, 734.  
 Halley's formula, 192.  
 — law, 589.  
 Hamilton-Jacobi differential equation, 126, 127, 130.  
 Hamilton's characteristic function, 128.  
 — equations, 121.  
 — Principle, 123, 124, 693.  
 Hankel functions, 817.  
 Harmonic analyser, 55 n.  
 — vibrations, 93.  
 Heat conduction, differential equation of 488.  
 — engine efficiency, 517.  
 — function, 509.  
 — of fusion, 539.  
 — of sublimation, 539.  
 — of vaporization, 536.  
 — specific, 487.  
 — theory of, 484.  
 Heaviside layer, 455, 858, Ex. 115.  
 Heisenberg, 705.  
 Heitler and London, 712.  
 Helmholtz-Lagrange equation, 405.  
 Helmholtz's vortex theorems, 205.  
 Herpolhode, 141, 142.  
 Herschel effect, 738.  
 Hertzian oscillator, 335.  
 — vibrator, 338.  
 Hess's law, 512.  
 Hittorf's relation, 420.  
 Hodograph, 81.  
 Holonomous-rheonomous system, 121.  
 Holonomous systems, 114.  
  
 Hooke's law, 171.  
 Huygens' formula, 376.  
 — principle, 368.  
 Hydrodynamical equations, 196.  
 Hydromechanics, 191.  
 Hydrostatics, 191.  
 Hyperbolic orbits, 702.  
 Hyperfine structure, 743.  
 Hysteresis loop, 461.  
  
 Image distance, 399.  
 — space, 397.  
 Images, electrical, 275.  
 Impedance, 318.  
 Impulse, 86.  
 Incidence, principle angle of, 357.  
 Index (of complex number), 50.  
 Indicator diagram, 516.  
 Induction, 264, 473.  
 — law, 311.  
 — magnetic, 298.  
 Inertia, 83.  
 — of energy, principle of, 258.  
 — product of, 149.  
 Inertial force, 231.  
 — frames, 229.  
 — mass, 84, 254, 259.  
 — resistance, 167.  
 Insulating media, 349.  
 Intensity of magnetization, 298.  
 — of wave, 724.  
 Interference fringes, 395, 410.  
 — instruments, 410.  
 Interferometer, Fabry-Pérot, 411, 412.  
 Internal forces, 108.  
 Ion, 418.  
 "Ion cloud", 550.  
 Ionization bursts, 765.  
 — photo-, 432.  
 — potential, 432, 649.  
 — thermal, 432.  
 Irrotational flow, 196, 198.  
 — region, 197.  
 Isobares, 759.  
 Isogyre, 375, 853, Ex. 98.  
 Isometric pressure coefficient, 493, 557.  
 Isothermal compressibility, 493.  
 — lines, 498.  
 — volume elasticity, 493.  
 Isotope of hydrogen, 743.  
 Isotopes, 745.  
 Isotopic mass, 765 n.  
 Isotropic substance, 171, 278.

- Jacobian function, 107.  
 Joule heat loss, 296.
- Kater's pendulum theory of, 151.  
 Kelvin temperature scale, 518.  
 Kepler ellipse, 656, 659.  
 Kepler's first law of planetary motion, 91.  
 — second law, 89.  
 — third law, 92.  
 Kerr effect, 864, Ex. 133.  
 Kinematics, 81.  
 Kinetic energy, 86, 112, 113.  
 — theory of matter, 563.  
 Kirchhoff's formula, 376, 378.  
 — laws, 294, 620.  
 Kutta-Joukowski lift formula, 208, 210.
- Lagrange's equations of the first kind, 113, 117.  
 — — of the second kind, 118, 120.  
 Lambert's law, 617.  
 Laminar flow, 214.  
 Langevin function, 463.  
 Laplace's equation, 71, 269.  
 Larmor frequency, 772.  
 — precession, 664.  
 Lattice constant, 574.  
 — defects in dielectric crystals, 735.  
 — energy, 576.  
 Laue-London theory, 466.  
 Law of areas, 89.  
 — of equipartition of energy, 595.  
 — of gravitation, 89, 90.  
 — of mass action, 539.  
 Le Châtelier's law, 542 n.  
 Left-handed elliptic motion, 59.  
 Legendre functions of second kind, 824.  
 — polynomials, 699, 820.  
 — — orthogonality of, 822.  
 Level surfaces, 21.  
 Light quantum, 432.  
 — principal velocities of, in crystal, 362.  
 Line integral, 22, 72.  
 — of nodes, 136, 137.  
 — spectra, 680.  
 Linearly polarized waves, 64.  
 Lines of force, 264.  
 — of motion, 194.  
 Liouville's theorem, 585.  
 Liquid films, 227.  
 Lissajous figure, 58, 60, 132.  
 Local time, 255.  
 Logarithmic decrement, 96, 100.
- Longitudinal mass, 476.  
 — waves, 64.  
 — — in bars, 180.  
 Lorentz contraction, 243, 244.  
 — transformation, 240, 249, 250.  
 — — consequences of, 242.  
 — triplet, 667.  
 Loschmidt's number. See *Avogadro's number*.
- $\mu$ -mesons, 779.  
 Macro-state, 627.  
 Magnetic domains, 462.  
 — lens, 431, Ex. 112.  
 — permeability, electron theory of, 450.  
 — reluctance, 308.  
 Magnetization by rotation, 464.  
 Magnetized shell, 303.  
 Magnetocaloric phenomena, 528.  
 Magnetomechanical parallelism, 464.  
 Magnetomotive force, 308.  
 Magneton, Bohr, 459, 665, 744.  
 — Weiss, 459.  
 Magnetostatic field, 297.  
 Magnification, angular, 401.  
 — in depth, 400.  
 — transverse, 399.  
 Magnus effect, 211.  
 Malus, law of, 397.  
 Manometer method, 290.  
 Mass action, law of, 422, 539, 542.  
 — correction, 765 n.  
 — defect, 765 n.  
 — number, 765 n.  
 — of photon, 688.  
 — spectrograph, 441, 442.  
 — variability of, 252.  
 matrix calculus, 46.  
 Mattauch's laws, 761.  
 Matter waves, 693.  
 Maxwell-Boltzmann law, 588.  
 Maxwell's equations, 311, 312, 467.  
 — relation, 331.  
 Mayer's equation, 505.  
 Mean anomaly, 137.  
 — free path, 567.  
 — relative velocity, 566.  
 Media, anisotropic, propagation in, 360.  
 Meissner and Ochsenfeld, 466.  
 Meitner, 778.  
 Melting-point curve, 534.  
 Membranes, 183.  
 Mesons, 778.  
 Metals, optics of, 357.



- Metals, properties of, 725.
- reflecting power of, 359.
- Michelson-Gale experiment, 478.
- -Morley experiment, 237, 477.
- stellar interferometer, 375, 853, Ex. 98.
- Micro-state, 627.
- Millikan oil-drop method, 425, 426.
- Mobility of an ion, 419.
- Modulated vibrations, 57.
- Modulus of compression, 173.
- of elasticity, adiabatic, 508.
- — isothermal, 508.
- Molecular diameter, 572.
- Moment of deviation, 155.
- of inertia, 146-9.
- — principal, 150.
- of torsion, 147, 297.
- Momentum, 84.
- Monogenic function, 69.
- Moseley's law, 655.
- Moving bodies, electrodynamics of, 471.
- Moving coil, induction in, 472.
- Multiple products, 15.
- Multiplets, theory of, 669.
- Multiply periodic systems, 128.
- Muscular action model, 805.
- Mutual inductance, 313.
- induction, 313.
- — coefficient of, 313.
- Nabla operator, 30.
- Nernst heat theorem, 553.
- Neumann function, 830, Ex. 26.
- Neutrino, 778.
- Neutron, 752, 757.
- mass of, 753.
- Newton (unit), 266.
- Newton's laws of motion, 84.
- third law, 108.
- Nicol prism, 374, 375.
- Non-conservative forces, 88.
- Non-divergent field, 302.
- Non-irrotational motion, 198 n.
- Normal acceleration, 82.
- surface, 368.
- Nuclear attraction, 764, 765.
- electrons, objections to, 757.
- isomerism, 763.
- magneton, 744.
- moment, 743, 744, 772.
- physics, 739.
- Nucleus, 431, 645, 646.
- artificial transformation of, 747.
- excitation of, 747.
- Nucleus, liquid-drop model of, 763.
- motion of, 651.
- Nutation, 154.
- cone of, 158.
- of force-free top, 158.
- Object distance, 399.
- space, 397.
- Oersted (unit), 298.
- Ohm's law, 293.
- — for electrolytes, 417.
- — for metals, 446, 447.
- Onnes, H. K., 466.
- Onsager theory of electrolytic conductivity, 421, 423.
- Operator  $\nabla$ , 30.
- Optical path, 395.
- systems, resolving power of, 409.
- Optics, electron, 443.
- geometrical, 395.
- interference, 395.
- Orbits, electronic, perturbation of, 662.
- hyperbolic, 702.
- Orthogonality, 189.
- Ortho-hydrogen, 862, Ex. 129.
- Orthotomic system, 264.
- Oscillations of plasma, 801.
- Oscillator, stabilization of, 794.
- Osculating plane, 19.
- Osmotic pressure, 547, 594.
- $\pi$ -mesons, 779.
- Packing fraction, 765.
- Parabolic velocity profile, 215.
- Para-electric susceptibility, 280, 450.
- Para-hydrogen, 862, Ex. 129.
- Paramagnetic, 450.
- Parelectric susceptibility, 450.
- Partial pressures, Dalton's law of, 496.
- Paschen-Back effect, 667, 743.
- Paschen series of spectral lines, 650.
- Paschen's law, 436.
- Pauli and Sommerfeld, electron theory of, 449.
- principle, 632, 679, 715.
- Pedal-surface, 370.
- Perihelion, 136, 137, 260.
- Periodic system, structure of, 674.
- systems, 128.
- Permeability, 298.
- magnetic, electron theory of, 450.
- Permittivity, 267.
- Perpetuum mobile, of first kind, 502.

- Perpetuum mobile, of second kind, 517, 519, 521, 859, Ex. 123.  
 Perturbation, 731.  
 — of electron orbits, 662.  
 Perturbations, theory of, 707.  
 Phase, 493.  
 — change of, in reflection, 355.  
 — difference, 51.  
 — integral, 130.  
 — orbit, 604.  
 — space, 584.  
 — velocity, 694.  
 Phosphorescence, 736.  
 Photo-electrons, 433.  
 Photon, 688.  
 Pickering series of spectral lines, 653.  
 Piezoelectric constants, 792.  
 — moduli, 793.  
 Piezoelectricity, 792.  
 Planck's law of radiation, 621, 622.  
 — quantum of action, 432.  
 Plane circulatory motion, 207.  
 — mathematical pendulum, 104.  
 — wave, 62.  
 Planetary motion, 90.  
 Plasma, oscillations of, 801.  
 Point charge, 275.  
 Poiseuille's law, 214, 215.  
 Poisson's equations, 268.  
 — ratio, 171.  
 Polarization, 280, 284, 350, 528.  
 — dielectric, 451, 473.  
 — electrical, 450.  
 — elliptical, in reflection, 355.  
 — linear, in reflection, 355.  
 — magnetic, 450.  
 — vector, 450.  
 Polhode, 141, 142.  
 Polygon of forces, 145.  
 Ponderomotive forces, 285.  
 Porous-plug experiment, 509.  
 Positive rays, 440.  
 Positron, 752, 754, 757.  
 Potential, 87, 88.  
 — barriers, 721.  
 — elastic, 175, 176.  
 — energy, 113, 114.  
 — excitation, 432, 649.  
 — flow, 198.  
 Gibbs, 525, 527, 532 n., 541.  
 — retarded, 340.  
 — thermodynamic, 524, 526.  
 Power factor, 319.  
 — of a lens, 403.  
 Poynting vector, 331, 332, 356.  
 Precession, 157.  
 — cone of, 158.  
 — Larmor, 664.  
 — of axis of earth, 158.  
 Pressure, radiation, 613.  
 Primary cosmic rays, 780.  
 Principal extensions, 167.  
 — normal, 19.  
 — plane, 399.  
 — point, 399.  
 — rays, 786.  
 — series, 660.  
 — stress, 169.  
 Principle of correspondence, 655.  
 — of uncertainty, 628 n.  
 Prism, Nicol, 374, 375.  
 — resolving power of, 856, Ex. 108.  
 — spectrograph, 409, Ex. 105, 855.  
 Product of inertia, 149.  
 Projectiles, deviation of, 839, Ex. 53.  
 Propagation in anisotropic media, 360.  
 Proper functions, 186.  
 — time, 255.  
 — values, 185.  
 — — of wave equation, 698.  
 Proton, 646, 757.  
 Pupil, entrance, 785.  
 — exit, 785.  
 Quantum, light, 432.  
 — of action, Planck's, 432.  
 — statistics, classical, 603.  
 "Quantum-mechanical square", 683.  
 Quartz vibrator, 797.  
 Quasi-elastic forces, 93.  
 Rabi, 772.  
 Radiation pressure, 613.  
 Radioactivity, 745.  
 — artificial, 756.  
 — induced, 756.  
 Radius of gyration, 148 n.  
 Raman effect, 689.  
 — — in wave mechanics, 716.  
 Raoult's law, 546.  
 Ray axes, 371.  
 — surface, 369, 370.  
 Rayleigh waves, 182.  
 Reactance, 317.  
 Rectifier, 434.  
 Reference frame, uniformly rotating, 232.

- Reflecting plane, 387.  
 — power, 354.  
 Reflection, optical law of, 349.  
 Refraction and anisotropic medium, 371.  
 — electron theory of index of, 450.  
 — external conical, 373.  
 — index of, 331.  
 — internal conical, 373.  
 — molecular, 455.  
 — optical index of, 451.  
 — optical law of, 349.  
 Refractivity, 455.  
 Relative aperture, 855, Ex. 107.  
 Relativistic mechanics, 228.  
 Relativity, generalized theory of, 258.  
 Reluctance, magnetic, 308.  
 Residue, 74.  
 Resistance, 293.  
 Resistivity, 294.  
 Resolution, limit of, 410.  
 Resolving power, 384, 409.  
 — — of prism, 856.  
 Resonance, 96, 99, 319.  
 — curve, 99, 102.  
 — energies, 754, 762.  
 Retarded potential, 340.  
 Reversible process, 503.  
 Revolving electrons, 463.  
 Reynolds's number, 215.  
 Richardson-Einstein-de Haas effect, 464.  
 Right-handed elliptic motion, 59.  
 "Ritz term", 659.  
 Röntgen and Eichenwald, experiments of, 475.  
 Rotating earth, free fall on, 231.  
 Rotation of a vector, 27 n.  
 Rotation-vibration bands, 682.  
 Rowland's experiment, 474.  
 Rutherford, 747.  
 Rydberg number, 651.  
 "Rydberg term", 659.  
  
 Sagnac's experiment, 478.  
 Sargent's formula, 747.  
 Saturation, 461.  
 — current, 433.  
 Scalar, 7.  
 — product, 10.  
 Scattering power, 719.  
 Schmidt limits, 774.  
 Schmidt-Schüler curves, 774.  
 Schrödinger, 705. —  
 — equation, 696, 722, 726, 727, 774.  
 Scintillation method, 740.  
  
 Scleronomous systems, 114.  
 Screw motion, 142.  
 Second law of thermodynamics, 514.  
 Secondary cosmic rays, 780.  
 Self-inductance, 314.  
 Self-induction, 313, 314.  
 — coefficient of, 314.  
 Shadow, geometric, 382.  
 Sharp series, 660.  
 Shear modulus, 173.  
 Shell integral, 14.  
 Signal velocity, 66.  
 Simple harmonic vibration, 50.  
 — pendulum, 103.  
 Simultaneity, 242.  
 Sine condition, 785.  
 Single-ray direction, 373.  
 Singular point, 73.  
 Skin effect, 324.  
 Smoke rings, 206.  
 Snell's law, 350.  
 Soap bubble, 227.  
 Solenoidal field, 289.  
 Solid state, 607.  
 Sound waves, 212.  
 Space curves, 17.  
 — relativistic conception of, 240.  
 Spark discharge, 433, 436.  
 — spectra, 674 n.  
 Specific heat, 503, 529, 595, 605.  
 — — of solids, Debye's theory, 608.  
 — — rotational, 861, Ex. 129.  
 — inductive capacity, 267 n., 275.  
 — resistance, 294.  
 Spectra, arc, 674.  
 — rotation, 702.  
 Spectral lines, 719.  
 — — intensity relations of, 703.  
 Spectrograph, prism, 409, 855.  
 Spectrum of alkalis, 657.  
 — of hydrogen, 652.  
 — of ionized helium, 652.  
 — order of, 384.  
 — X-ray, 653.  
 "Sphere of influence", 566.  
 Spherical condenser, 274.  
 — harmonics, 699, 819.  
 — — relations between, 820.  
 — wave, 63.  
 Spin of electron, 669.  
 Spontaneous magnetization, 462.  
 Standing waves, 67.  
 Stark effect, 475, 662.  
 Starting vortex, 211.



- Statistics, Fermi-Dirac, 625, 631.
- quantum, classical, 603.
- Stefan-Boltzmann law, 624.
- Steiner's theorem, 148.
- Stern and Gerlach, experiments of, 665.
- "Stokes" line, 721.
- Stokes resistance, 418, 419, 425.
- Stokes's law, 218.
- — of fluorescence, 721.
- theorem, 25, 72, 203, 204.
- Strain, 162.
- ellipsoid, 171.
- tensor, 171.
- Stream function, 200.
- lines, 194, 201.
- Stress, 167.
- tensors, 171.
- Strings, transverse vibration of, 183.
- Subordinate series, first, 660.
- — second, 660.
- Superconductivity, 466-70, 635.
- Superposition of plane waves, 68.
- Surface brightness, 614.
- curl, 345.
- invariant, 404.
- tension, 222, 223.
- waves, 182.
- zonal harmonics, 699, 820.
- Surfaces of discontinuity, 43.
- Susceptibility, 280, 450, 453.
- diamagnetic, 463.
- ferromagnetic, 458.
- paramagnetic, 458, 669.
- piezoelectric, 280, 455.
- Swinne formula, 747.
- Symbols in nuclear physics, 748 n.
- Systems of particles, 108.
- Tangent vector, 18.
- Tangential acceleration, 82.
- Taylor's series, 76.
- Temperature, critical, 499.
- scale, thermodynamic, 518.
- Tensor analysis, fundamental principles of, 32.
- antisymmetric, 36.
- ellipsoid, 37.
- invariants of, 172.
- symmetric, 36.
- Thermal radiation, 613.
- Thermionic tube, characteristic curve, 799.
- Thermochemistry, 511.
- Thermodynamic equilibrium, 547.
- Thermodynamic potential, 524, 526.
- Thermodynamics, first law of, 501.
- second law of, 514.
- third law of, 553.
- Thomson's theorem, 204, 205, 211.
- Timbre, 187.
- Time, relativistic conception of, 240.
- Top, spherical, 152.
- symmetrical, 153.
- theory of the, 151.
- Torricelli's theorem, 841, Ex. 60.
- Torsion, 20.
- Total heat, 509.
- reflection, 354.
- Transport numbers, 420.
- Transverse mass, 476.
- waves, 64.
- Trouton and Noble, experiment of, 475.
- "Tunnel effect", 747.
- Turbulence, 215, 216.
- Turbulent motion, 198 n.
- Unbounded elastic media, waves in, 180.
- Uncertainty principle, 628 n.
- Undulatory propagation, 212.
- Unital stress, 167.
- Unit plane, 399.
- vector, 9.
- Units, C.G.S., 266, 330.
- M.K.S., 266, 330.
- Uranium 1, 721.
- Valve, electrical, 434.
- Van der Waals' equation, 498.
- Van't Hoff's law, 547.
- Vapour pressure curve, 534.
- Variation, 78.
- Variations, calculus of, 75.
- Vector, components of, 8.
- concept of, 7.
- diagram, 52.
- differentiation of, 17.
- gradient, 36.
- multiplication by scalar, 8.
- polarization, 450.
- potential, 302.
- product, 12.
- Vectors, addition and subtraction of, 8.
- axial, 15.
- equality of, 8.
- polar, 15.
- Velocity, ray, 364.
- resonance, 99 n.
- vector, 81.

- Vibration, plane of, 351.
- Virtual displacement, 114.
- displacements, principle of, 113.
- Viscosity, 568, 571.
- coefficient of, 214.
- Viscous fluids, 214.
- Vortex filament, 205, 206.
- lines, 205.
- motion, 203.
- ring, 206.
- strength, 205, 206.
- tube, 205, 206.
  
- Wattless current, 319.
- Wave equation, 328.
- — proper values, 698.
- normal, 371.
- surface, 368, 369, 370.
- velocity, 61 n., 245.
- Wave-length, 62.
- "Wave-length, De Broglie", 695.
- Waves, 61.
- Waves, elliptically polarized, 64.
- linearly polarized, 64.
- Weiss law, 529.
- Wiedemann and Franz, law of, 447, 449.
- Wien's displacement law, 622.
- experiment, 474.
- Wilson's cloud chamber, 642.
- Wireless telegraphy, 341.
- Work, 85.
- Work of escape, 431.
- World-tensor, antisymmetric, 251.
- World vectors, 247, 250.
  
- Young's modulus, 171.
- Yukawa's theory, 778.
  
- Zeeman effect, 663, 671, 742.
- — anomalous, 667.
- — longitudinal, 863, Ex. 133.
- Zone lens, 394.
- plate, 390.

